

**U.S. ENVIRONMENTAL PROTECTION AGENCY**

**EPA SCIENCE ADVISORY BOARD**

**CLEAN AIR SCIENTIFIC ADVISORY COMMITTEE (CASAC)**

**PARTICULATE MATTER REVIEW PANEL**

**July 18, 2002**

**MR. FLAAK:** I would like to welcome everybody to today's meeting. I am Bob Flaak. I am the Designated Federal Officer for the Clean Air Scientific Advisory Committee. Those of you that have worked in this arena for a while certainly have probably seen me doing this once or twice before.

I have a couple of announcements to make before we get started, and let me go over those briefly. There are materials outside the room at the sign-in table outside. Please make sure that you pick up a copy of the agenda if you haven't done so already. We have a fairly busy schedule. Most of today is devoted to presentations by various individuals, both invited presentations which will take quite place principally this morning and into the early part of this afternoon, followed by public comments that will take place beginning around the middle of the afternoon until the end of today.

For rest rooms, they are located outside these doors on my right side, on many of your left sides. On the other side is a large staircase that comes down, and next to the cafeteria, past the phone banks, are the rest rooms.

If you come in and out of the room during the course of the meeting, I ask that you use the doors on the sides and not the ones up here in the front.

Dr. Frank Speizer, a member of CASAC, is not with us today. He is out of town. However, he has joined us on the telephone, and he will be with us through this morning's session and, Frank, I believe through tomorrow morning's session as well?

**DR. SPEIZER:** Yes.

**MR. FLAAK:** Okay, good. Thank you. Can you hear me okay?

**DR. SPEIZER:** Yes.

**MR. FLAAK:** Great, thank you.

I would just like to make a couple of announcements regarding the panel. There is information on the side table, I believe, or outside on the introductory information on the panel members. We will not go through an oral disclosure at this meeting, since we have the written materials.

I might point out that, under the conflict of interest rules of the Federal Government, as the Designated Federal Officer for this committee, I have evaluated the confidential financial forms of all of the members of this panel and have determined that there are no conflicts of interest that exist. Where appropriate waivers have been granted, in some cases, where potential conflict might be identified on individual holdings, those do not pose a

problem, and we have evaluated those in consultation with our attorneys at EPA, and we have determined no conflicts do exist.

Do any members of the panel have any questions this morning before I turn it over to Phil to get started?

(No response.)

**MR. FLAAK:** Okay. Les, you have all your people here?

**DR GRANT:** Mm-hmm (indicating affirmatively).

**MR. FLAAK:** Okay.

**DR. HOPKE:** This is the next in the reviews of the Criteria Document for Airborne Particulate Matter. We began this with a very preliminary draft a while back and then, last year, about this time, looked at a more detailed draft, provided significant comments back to the Agency on them. In the meantime, there has also been significant amounts of additional information, and this was compiled into the April draft version that you have seen.

As a result of some other discoveries that have occurred, then, this spring, a number of issues have arisen with regards to some of the statistical analyses used to try and identify the relationship between exposure and effects of airborne particulate matter, and this has, then, led to us needing to take a couple of steps backwards and carefully review these issues, see what they imply with regards to the

statistical evidence for relationships between exposure to particulate matter and adverse health effects and to then look at how we will then move to finalize this document to the point where we could close on it and move on in the process.

So, our primary purpose for today's session and the beginning of tomorrow morning is to carefully review the issues that have arisen and provide advice back to the Agency as to what sensible approaches could be taken to try and come to a best understanding of where we are with regards to the effects of the statistical problems and what really are the underlying relationships that we need to understand as we try and evaluate this Criteria Document.

So, as you can see, we are going to have this extensive set of presentations. I think it would be useful for us to start off by allowing everybody in the audience to know who we are. So, let me introduce myself. I am Philip Hopke. I am the A.R.D. Clarkson Distinguished Professor of Chemical Engineering and Chemistry at Clarkson University, and my background is mostly in data analysis and receptor modeling and some field studies in nucleation.

So, I would suggest we might go a quick round, starting over with Warren and...

**DR. WHITE:** I am Warren White from Washington University in St. Louis. My background is mathematics and aerosol science.

**DR. LIPPMANN:** Mort Lippman, New York

University School of Medicine, Environmental Health Science.

**DR. LIOY:** Paul Lioy, Environmental and Occupational Health Sciences Institute in New Jersey, exposure and environmental health sciences.

**DR. LEGGE:** I am Allan Legge with Biosphere Solutions in Calgary, Alberta in Canada, and I deal with issues related to air quality and environmental effects.

**DR. OBERDORSTER:** I am Gunter Oberdorster, University of Rochester, Department of Environmental Medicine. I am interested in inhalation toxicology of particles, non-fibrous and fibrous.

**DR. MCCLELLAN:** I am Roger McClellan, an independent advisor on inhalation toxicology and human health risk analysis in Albuquerque, New Mexico.

**DR. KOENIG:** Jane Koenig. I am at the University of Washington in Seattle. I am interested in physiological...human physiological responses to air pollution.

**DR. MAUDERLY:** Joe Mauderly with the Lovelace Respiratory Research Institute in Albuquerque, and I am interested in toxicology and dosimetry.

**MR. POIROT:** I am Rich Poirot. I am an environmental analyst with the State of Vermont, Department of Environmental Conservation, and I am interested in aerosol measurement methods and source apportionment techniques.

**DR. MILLER:** Fred Miller with CIIT

Centers for Health Research. My interest is dosimetry and toxicology. I am trained as a biostatistician.

**DR. VEDAL:** Sverre Vedal. I am a professor of medicine at National Jewish Medical and Research Center in Denver, Colorado. I am a chest physician and an epidemiologist.

**DR. ZIELINSKA:** I am Barbara Zielinska. I am a research professor at the Desert Research Institute in Reno, Nevada, and my interest is atmospheric chemistry, organic compounds for gas phase and particle entity.

**DR. WOLFF:** George Wolff from General Motors. I am an atmospheric scientist.

**DR. SAMET:** Jon Samet, Johns Hopkins Bloomberg School of Public Health, Department of Epidemiology, epidemiology and pulmonary medicine.

**MR. WHITE:** Ron White, National Osteoporosis Foundation and volunteer with the American Lung Association, and my interest is in air pollution and human health effects.

**DR. TAYLOR:** I am George Taylor. I am with the School of Computational Sciences at George Mason University in Fairfax, Virginia, and I am an ecologist, air quality specialist.

**DR. KOUTRAKIS:** Petros Koutrakis, Harvard University. I am in atmospheric sciences and an exposure assessment specialist.

**DR. HOPKE:** Thank you very much. Frank, you want to introduce yourself, too?

**DR. SPEIZER:** Yes, Frank Speizer, professor of medicine and environmental science at Harvard University.

**DR. HOPKE:** Thank you. So, again, the idea is today, we are going to focus on these questions underlying the epidemiology. Tomorrow, we will attempt to go through the other chapters, chapter by chapter.

Now, because of the limited time we have, we may not be able to complete the discussion of each chapter. We are going to have to set aside time to make sure that we get to all chapters, but what I would like to ask the panel members is that, in preparation for tomorrow, please look over your comments and look to the ones which highlight any of the major science problems so that we get those on the table. All of the minor technical things, editing, things like that, we can deal with in terms of providing the written comments to Dr. Grant and his staff who can take care of those things, but any issues you have with the way...with any of the science issues as they are presented in those chapters we would like to get on the table and, hopefully, be able to provide Dr. Grant and his team with some clear advice on where to go so that when we get together the next time, we really will be in a position to close on this document.

They have made a significant effort in revising

things with regards to our comments from last time. From what comments we have seen so far, we thought we might be able to actually close this time, but then these other issues have come up. I don't want you to feel that we are limiting the debate on the other chapters. We will have more time at the future meetings to do that, but let's make sure that we bring up any major issues that still remain in these chapters so that the problems are clearly outlined to the team that needs to revise them so that when we do get the epidemiology issues resolved and rewritten, we'll be at a point where everything else has really come together as well.

**MR. FLAAK:** I have one more comment. Let me ask, also, that the individuals on the panel who have provided me with their preliminary written comments, that information is contained in a composite document which is available on the table outside. For those of you that have yet to give me your comments, I have received written comments today from Allan Legge which we will get copied. I have received electronic copies of Dr. Hopke and Dr. Oberdorster's comments. If anyone else has comments to provide to me today, such as Dr. Koutrakis, thank you, and anyone else, please get those to me, and we will get those copied and make them available probably by the end of today.

For members of the public who are presenting today and who have written comments, I ask that you do not put the comments on the table outside but that you give those to me directly or to my assistant, Zisa. Zisa, I just saw you walk



in the room. Where are you? Zisa is at the back of the room. Either give them to her or give them to me so we can make sure that they get distributed to everyone appropriately, and, if necessary, we will get extra copies made.

**DR. HOPKE:** Thank you. Les, you are up.

**DR. GRANT:** Dr. Les Grant, director of the National Center for Environmental Assessment Division in Research Triangle Park and, obviously, head of the group responsible for preparation of the Criteria Document.

Well, nice to see you all again, Phil and other members of the PM review panel. I am joined here today to my right by William Wilson who is our PM team leader and other members of the PM team on my staff responsible for helping to prepare the document as well as several of the consultants that aided us in doing so.

Well, where to start? Perhaps with a few words to help place this meeting in context. In case folks are not aware of it, today is the fifth anniversary of the new PM standards being promulgated back on July 18th in 1997. Lots of things have happened since then. Perhaps a couple highlights might be useful.

First of all, to note that, at the time of that promulgation and in light of a lot of controversy regarding some of the scientific bases and so forth, the Congress appropriated \$50 million a year to expand the PM research program within EPA. Also, there was the establishment of a

National Research Council or NRC Committee on PM Research Priorities to help oversee and advise EPA management in conduct of the research program.

I think one of the key things here is the fact that, indeed, that program developed very rapidly, quite extensive research both in terms of our intramural Office of Research and Development PM program, increased funding for STAR grants and Center grants and so forth, the establishment also, separately from the program, of the National PM<sub>2.5</sub> monitoring network and speciation network and so on to work jointly with some of the expanded research efforts.

One of the consequences of this is a truly unprecedented outpouring of new research findings through the last number of years, and as you all know, the rapid rate of publication of these research findings certainly has posed for all of us quite a huge challenge in trying to capture and interpret or assess whatever the essence of what that research all means.

The CASAC comments and public comments on our two previous external review drafts that Phil mentioned back in '99 and 2001 have been extremely helpful as we have gone on to try, you know, to come up with further revision of the document. We have taken the comments seriously, believe me, and I think we've made some very real progress in what is contained in the third external review draft, not that there isn't still some additional progress that needs to be made, obviously, but we think we have been able to make some, you

know, very major improvements and advance in what has been assessed there.

I would just as far as a quick summary overview of some of the key revisions, a number of the revisions made and some of the key findings are contained in a handout that were placed at each of the CASAC members' places. It is also available out here out front, and I presume that everybody has picked it up. I would like to just run through very quickly and highlight a few of the things here in the handout.

First off, we have added the executive summary. It now appears at the outset of the document before the ensuing chapters, and, essentially, as its name intends, it summarizes key points and conclusions from all the ensuing chapters.

The introduction to the document in Chapter 1 has been revised to explain as a key change, the general flow of the rest of the document, and that is to follow the risk assessment paradigm as per the framework that the NRC Committee on PM Research Priorities has employed in some of their work.

So, it goes from Chapter 1, the introduction, to Chapter 2 which provides background information on physics, chemistry, and measurement of PM to 3 which talks about sources of emissions, air concentration in the U.S., and then, after a Chapter 4 discussion of environmental effects, we go through a sequential discussion of information on human

exposure, dosimetry, toxicology, epidemiology, and, lastly, we wrap up with Chapter 9 as an integrated synthesis, especially of the air quality aspects and health-related things.

The next is an augmented discussion in Chapter 2 of the chemistry and physics of atmospheric PM, and this continues to support separating PM<sub>10</sub> into fine and coarse thoracic fractions, and it provides quite of bit of additional information, new information or whatever on ultrafine and inhalation mode PM within the fine fraction. There is more information in the measurement section of that chapter on the FRM for PM<sub>2.5</sub> and improved ways to provide a sharp cut at the 2.5 diameter.

We also talk about so-called difference technique and the dichotomous sampler as candidates for coarse fractions PM, that is, PM<sub>10-2.5</sub> sampling, and we have added quite a bit of discussion with regard to progress made in sampling and analysis of semi-volatile components, including ammonium nitrate, semi-volatile organic compounds, and so forth, as well as some discussion about progress made in continuous monitoring for PM mass and several different PM components.

Chapter 3, the one on concentration sources and emissions of atmospheric PM, discusses primary and secondary sources of fine and coarse PM and concentrations of PM<sub>2.5</sub> and, to some extent, PM<sub>10-2.5</sub> derived from the new National

Monitoring Network are analyzed and reported in terms of both of spatial distribution on national and urban scales and temporal distribution, that is, our seasonal and diurnal time scales.

Very importantly, everybody needs to keep in mind that we are looking at what is just now emerging from a couple years of data coming out of this national network, and the analyses of spatial variability of PM<sub>2.5</sub> concentrations which were done using about 27 U.S. urban areas do indicate that there is quite a bit of...or a varying degree of heterogeneity in PM<sub>2.5</sub> levels, and note is made cautioning in using data from the AIRS data base, then, to characterize community exposures to PM<sub>2.5</sub>, and these have some implications, perhaps, for some of the epidemiology analyses that are talked about in Chapter 8.

There were similar analyses done or added for the coarse fraction PM, PM<sub>10-2.5</sub>, as per recommendations from CASAC during the last go-around. These are pretty limited compared to the PM<sub>2.5</sub> evaluation, and some of the reasons I have listed in here, very importantly in many areas, PM<sub>10</sub> and PM<sub>2.5</sub> are not necessarily measured at the same sites.

Secondly, in many areas, there is little overlap that ties between the sampling schedules, so you have limited opportunities of having the same day on which you have PM<sub>10</sub> measurements and PM<sub>2.5</sub> at the same site that you can then

subtract  $PM_{2.5}$  from  $PM_{10}$ .

Lastly, any errors that you have in terms of measurement for  $PM_{10}$  or  $PM_{2.5}$  do propagate, then, into the  $PM_{10-2.5}$  concentration estimates. Sometimes, these yield negative values for  $PM_{10-2.5}$ .

In any event, we have tried to do about the best we could with the available data with regard to trying to characterize the coarse fractions distributions and so on.

We do plan to add some further similar analyses expanded to take into account more data collected over another six months or so in the same or additional cities as they are added to the available AIRS data base. Secondly, we have in mind to add analyses of data available from the Los Angeles Basin from dichotomous samples which give you both  $PM_{10}$  and  $PM_{2.5}$ , obviously, at the same location, same sampler, and compare that against what you get when you use the difference method where you subtract  $PM_{2.5}$  from  $PM_{10}$  derived from different monitors.

There is a great deal of information that we have presented, especially in the appendices and so on, to Chapter 3 on chemical composition data that are being derived from our new speciation network and statistical techniques for determining source categories and the contribution of PM mass due to specific source categories are described in the chapter as well.

The question for epidemiology of how well a community monitor can represent an entire urban area is, in part, assessed by examining the homogeneity or heterogeneity among levels of PM<sub>2.5</sub>, PM<sub>10</sub>, PM chemical components, and source category contributions across U.S. areas in this chapter.

Chapter 4 on the environmental effects of PM, the vegetation and ecosystems section has been augmented by quite an extensive new discussion, again, as per CASAC recommendation last time, of key determinants of wet and dry deposition of particles on plants and ecosystems. We have also redrawn quite a number of the figures to enhance the clarity of presentation of them in the chapter, and we have added a discussion, again, as per recommendation of CASAC, of limited available information on urban ecosystems.

The visibility section has been revised to have a clearer discussion, perhaps, of methods for measuring PM visibility effects and also to talk a little bit about or summarize information on the visibility monitoring programs being carried out by EPA and some of the other Federal agencies.

As for the climate change section at the end of Chapter 4, we have shortened that. We have deleted out the associated appendices. Those appendices were in there mainly as a means to sort of scanning in sort of executive summary information from other extensive assessments that were just then in, what, March or April of 2001 not yet published but

available to us. Now, these are published, and we make reference to them and so forth but don't feel the need to have to have sort of the executive summary materials in the appendices for the chapter.

Turning to Chapter 5 on human exposure to PM, it has been revamped to better explain different components of personal exposure, and it describes techniques for determining ambient and indoor-generated components of personal exposure, and information, new information, is presented on size, composition, and sources of personal exposure.

The second major thing, studies of ambient concentrations and personal exposures of cohorts provide information on relationships between ambient, non-ambient, and personal exposures to PM both for healthy and susceptible populations. Some of the new analyses of the PTEAM data base show that the ambient PM concentration is highly correlated with personal exposure to ambient PM, but the correlation between ambient concentration and non-ambient exposure is very low, near zero.

Also, analysis of exposure error suggests that non-ambient exposures will not bias epi studies of statistical relationships between ambient concentrations and health outcomes, but the difference between ambient PM concentrations and ambient PM exposures can bias the relative risk per unit of ambient PM concentration. Reductions in the community average ratio of ambient PM exposure to ambient PM



concentration with increasing use of air conditioning does appear to explain at least part of the homogeneity...heterogeneity found in multi-city epi studies.

The newly added exposure studies, certain of them, also suggest that gaseous pollutants such as NO, CO<sub>2</sub>, SO<sub>2</sub>, or NO<sub>2</sub>, rather, and ozone are, at times, or can be surrogates, not necessarily confounders, of PM<sub>2.5</sub>, given new findings that indicate or suggest that ambient concentrations of the gases are poorly correlated with personal exposures to the gases but are well correlated, at times, with ambient PM<sub>2.5</sub> concentration.

As for the dosimetry of particulate matter, the chapter has been extensively revised to more clearly and, I think, more transparency, shall we say, elucidate, then, human respiratory tract deposition, clearance, and retention of particles of varying size or chemical and physical composition in the three major compartments of the respiratory system, the extrathoracic, tracheobronchial, and alveolar.

We have added in quite a bit of new graphics as per suggestions from the committee last time. I think these are quite helpful. They elucidate further the deposition and retention patterns and so on.

Revisions have also been made regarding our improved understanding of species difference in deposition, translocation, and clearance of particles, and, hopefully,

this may be able or provide us with a better possibility to extrapolate and interpret toxicologic data obtained from laboratory animal studies.

Expanded discussion has also been accorded to more sophisticated and versatile mathematical and fluid computation models of respiratory tract particle deposition, clearance, and retention, and these are particularly helpful in identifying factors that may increase susceptibility to PM exposure and also delivered or retained dose to lung tissue and then perhaps the associated ambient PM health effects. This includes looking at differences in deposition, for example, as they relate to age, to preexisting disease states, and so on.

Chapter 7 on the toxicology of PM in human and laboratories animals has been revamped. It is aimed to answer several questions that are posed at the outset of the discussion, I think, such as what are the types of toxicologic effects caused by exposures to PM and including, especially, relevant ambient air concentration as a key point; what characteristics, size, or composition of PM contribute to any of the observed toxicity; what are combined effects of PM and other pollutants; what mechanism may be involved in the toxicologic responses to PM; and what factors affect individual or subpopulation susceptibility to the effects of PM.

The emphasis is placed in the Chapter 7 discussions, in the revised discussions, on the assessment of

new data both from controlled studies of particles collected from emission sources and ambient samplers such as impactors and diffusion denuders, and then, secondly, data obtained by use of aerosol concentrators that provide a technique for exposing humans or laboratory animals by inhalation to concentrated ambient particles or CAPs.

The new studies pretty consistently tend to indicate that the biologic effects of PM may be a function not only of particle mass deposition but of some other characteristics, particle numbers, total surface area of the particles, particle acidity or surface chemistry, charge, and composition of the particle in addition to other exposure variables and a number of other environmental and host factors.

New studies also provide quite a bit of important additional information regarding potential mechanisms of action of PM or PM constituents. We think that, very importantly, the additional information that is brought in here and expanded on neurological mechanisms are likely or perhaps, whatever, suggestive of contributing to cardiovascular effects is some pretty...you know, one of the more important things that has been added. There are quite a number of other new pieces of information as well on other types of mechanisms.

New information from the laboratory animals studies also provide plausible insight into risk factors enhancing the toxicologic properties of PM, these being useful in

identifying potential susceptible human population groups.

We think, collectively, that these important new findings assessed in Chapter 7 do add quite a bit, substantially, we think, to the weight of evidence that argues for the plausibility of ambient PM exposure effects on humans and, especially, in comparison to what we have available in the way of very limited information in the 1996 PM Criteria Document.

Turning to the epidemiology of human health effects of ambient PM, we find that we have done quite a bit to revise the introductory background information that includes, we think, quite improved discussion and illustrative figures on confounding and effect modification concepts. This is in keeping, again, with recommendations from the committee to try to make a clearer and more transparent introduction to these concepts.

Also, section 8.1.3 provides information on approaches used to acquire the published studies for consideration in the chapter and rationale for relating PM excess risks to standardized PM increments. There had been some discussion last time perhaps we should go to some standard single increment, say,  $10 \text{ } \mu\text{g} / \text{m}^3$  or whatever, and when we went back in and had a look at the actual data from the monitoring networks, it did appear that it still is probably substantially or whatever justified to go ahead and continue to use the  $50 \text{ } \mu\text{g} / \text{m}^3$  increment for  $\text{PM}_{10}$  and then 25

for PM<sub>2.5</sub> and the PM<sub>10-2.5</sub> fractions, giving the range of kind of ambient exposures.

In revising the ensuing sections on mortality and morbidity, we included more than 50 new studies published in 2001 and another 10 in 2002. That is in addition to the more than 200, maybe 250 studies, that have appeared since 1996 and were discussed in the second External Review Draft of the PM Criteria Document. So, as you can see, there continues to be quite a rapid additional increments or whatever to the epidemiology data base which certainly has made it pretty difficult for us to keep up with it and to try to capture it and present it succinctly in this document.

What we have done, in keeping with recommendations from the committee last go-around on the second External Review Draft, is to shift some of the detailed summary materials, tables, whatever, from the main body as they appeared in the second External Review Draft back into the appendices for Chapter 8 in this third draft. There are tables arranged by exposure duration, short- or long- term exposure, PM indicator such as PM<sub>10</sub>, PM<sub>2.5</sub>, PM<sub>10-2.5</sub>, and then different health endpoints, mortality versus different morbidity indices, breaking out morbidity indices, for example, of hospital admissions, physicians' visits, physiological or whatever pulmonary function changes and so on.

We have added more concise summary tables to the

main text and, again, organized by health outcomes and present some key features of the studies. We try also to bring forth in many of those tables quantitative or information succinctly provided on quantitative increases in risk per PM increments observed in models that have only PM alone in them or with one or more copollutants. It gets pretty tough to pull out key information from these studies in which quite a number or variety of PM models, you know, PM alone or with some various copollutants are in there.

Some of the more notable published epidemiology findings include, as follows:

First, effects of long-term exposure to  $PM_{2.5}$  on mortality appear to be confirmed rather well by published reanalyses of the Harvard Six-Cities Study and American Cancer Study. These are the prospective chronic exposure studies. These reanalyses by HEI-supported investigators and then, secondly, also, by the recently published extension of the ACS study to include a longer exposure period or longer follow-up period, I guess, would be more accurate. The ACS study, recently published in early 2002, also provides probably the strongest evidence yet of increased lung cancer risk being associated with long-term fine PM exposure.

We also have a number of new time-series studies that have looked at comparing  $PM_{2.5}$  and  $PM_{10-2.5}$ , the fine and coarse fractions of  $PM_{10}$  in the same study, and those appear to confirm the effects of  $PM_{2.5}$  overall and suggest

possible effects of  $PM_{10-2.5}$ , and, generally, they indicate higher risk and higher statistical significance for either the fine or the coarse fraction compared to  $PM_{10}$ .

There is also discussed intervention studies, if you will, or natural experiment studies, however you wish to term these. These are situations in which there is a change in PM levels, fairly dramatic changes or whatever, in which they are reduced, perhaps due to shutdowns of certain industrial operations, as in the case of the steel mill in Utah Valley, or to shifts in traffic patterns or movement of subjects to areas with different pollution levels.

Basically, these studies tend to show that at a decreased PM exposures due to these factors I just mentioned, you also get corresponding decreases in health endpoints such as mortality or morbidity. Conversely, the opposite. If you kick back in and start up, say, the steel mill in Utah or whatever, you see again increases in the health endpoints corresponding to that.

The epidemiologic analyses, we think, have identified relationships between mortality and morbidity and a number of different specific PM characteristics, PM physical and or, for example, size or chemical components, also in terms of associations with different source categories. As an example, some of the factor analysis studies, then, showing combustion-derived particles having strong relationships versus not very much coming out

indicating associations with the health effects with crustal coarse mode particles.

We have added quite a bit of discussion or revamped the discussion in section 8.4 of Chapter 8, and that modified, expanded discussion, for example, includes further discussion of the potential importance of intra-urban spatial differences in concentrations of PM and other potentially confounding pollutants as a possible source of measurement error. This is pulling some information out of Chapter 3.

We also discuss relationships among concentrations in air monitors, outside and indoor microenvironments, and personal exposure in relation to measurement error and confounding based on some of the discussions in Chapter 5.

It is also made the effects of local and regional variations in fine and coarse particle concentrations, their sources and composition as it may be relevant to epidemiology.

We have also highlighted in a section there lung cancer effects associated with exposure to ambient particles, both talking a little bit about some of the historic information that was available in the past and then the much better evidence or convincing evidence that is derived from the more recent prospective cohort studies and, especially, the ACS study, and we make cross reference over to the likely importance of diesel PM as being a likely important contributor as per our separate EPA Diesel Document that has just been published and now approved as final in 2002.



Obviously, as I think by now pretty much everybody knows and Phil alluded to earlier, some of the latest developments coming along with regard to certain statistical issues preclude a full or complete review of Chapter 8 epi assessments and the associated valuations of epidemiology findings in Chapter 9, the integrative synthesis. This is due to some recently surfaced statistical issues that potentially affect quite a number of PM time-series analyses that we assess in Chapter 8.

One such issue is the convergence issue, so-called, that relates to one aspect of certain software or whatever used to fit generalized additive models, GAM models or whatever, and, more specifically, the use of certain preset default criteria which terminate iterative curve-fitting routines or subroutines in some of the software.

Another issue concerns variance estimation in GAM modeling which, under certain circumstances using various software, can lead to underestimation of standard errors and confidence intervals by which levels of statistical significance are judged for possible PM-related mortality and morbidity effects.

Some of these issues were brought to our attention shortly after our release of our third External Review Draft, brought to our attention by HEI, and we did distribute out to CASAC and to all the recipients that we could identify as such of that third External Review Draft a letter that essentially highlighted the fact that these issues have been

brought to our attention and that we, EPA, were proceeding to address them but that it was recognized we would not be able to fully resolve and arrive at a complete review of the epidemiology information at this meeting. Rather, you know, we need to focus on discussion of these issues, and, indeed, in light of the importance of these issues, we have arranged for invited presentations by several different statistical and epidemiologic experts who have been involved in evaluating them and looking at some of their implications.

These presentations that are scheduled on the agenda during the next couple hours or so, we hope, along with ensuing discussions, will elucidate very important information on the nature of the statistical issues, some of their potential implications for evaluating the published time-series analyses, and we hope to hear about some of the preliminary results or reanalyses of some of the major PM epidemiology studies.

Attached to this handout in Attachment A is a set of information on some of the efforts carried out by EPA in trying to identify which of the various studies that were discussed in the '96 document and then also in this third External Review Draft, which of the many, about 350 of these studies, may have been affected by some of these statistical issues. I think it is important to note that with regard to those studies cited in the '96 document, relatively few of them were affected by or potentially affected by these issues dealing, especially, with GAM, general additive model or

whatever, analyses, whereas the newer studies, quite a number, more than 50 percent of some categories of these studies cited in the '96...excuse me...cited in the current third External Review Draft may be affected.

Quite a bit of effort has gone in, as you see in the table, effort put in by our associates over in OAQPS and also our staff in having a look at published descriptions of the methods and listing them out here in these tables as per the published descriptions, some contact with some individual investigators to get further clarifications, but as you can see, it is quite a big job if one is going to go through, you know, each of them, but efforts are being made to clarify and get additional information beyond what was in the published papers with regard to some of these different studies, and then to, hopefully, through discussions today and our further evaluations, to be able to arrive at what next in the way of dealing with them.

We still think CASAC review of Chapter 8 would be very valuable to us and help us in our next revision of that chapter, and that includes having CASAC comment on the adequacy, soundness, transparency of some of the newly-added revisions, discussing the fundamental concepts related to confounding and effects modification, also the summarization of key outcomes of mortality and morbidity studies in the concise summary tables and figures presented in the Chapter 8 main body, and our shifting of the lengthy descriptive summary tables to the appendices.

Comments on the adequacy and soundness of the text discussions of various important studies would be helpful as well, and also on the interpretive evaluation of epi findings in Section 8.4. That includes assessment confounded by copollutants, some pretty extensive discussions of figures related to that, the roles of particulate or specific particulate matter components, discussions of heterogeneity of PM effects, and so on. And it would also be helpful, obviously, to have comments on the cogency or soundness and completeness of our key conclusions presented in Section 8.5.

As for the integrative synthesis, we have added that synthesis now for the first time in this third External Review Draft. We have tried to organize it along the flow of the basis risk assessment paradigm framework, similar to what was used across the earlier chapters, and it is structured in a manner to address issues of the type that were posed by the NRC PM Committee on PM Research Priorities as topics 1 through 10 in their several published reports during the past several years.

Just a couple things of note as far as some of the health-related findings or whatever just to highlight it as far as the integrative synthesis, first, that is to note that we now do discuss and bring into play new information that we think that is highly suggestive of new mechanisms having been discovered whereby a variety of PM characteristics or components may cause biological responses. This includes progress made in understanding biological mechanisms whereby

deposition of PM in the lung can cause effects on the heart and blood that may lead to sudden death.

The table down below, Table 1, illustrates some of the types of particle characteristics, if you will, that have been associated with health effects either in toxicology or epidemiology studies, as discussed earlier in the chapters and so forth.

And we believe that the combination of the toxicology and epi studies suggest that different chemical components and physical characteristics of PM from different source categories may have qualitatively and quantitatively different effects. We think, as of now, you really cannot rule out the possibility of just about any given PM<sub>1</sub> 0 fraction or chemical constituent having perhaps toxic potential at ambient or near-ambient concentrations.

There may be a few exceptions to that, as I mentioned earlier. Not too much evidence coming from the factor analysis studies of these kinds of effects being associated with crustal materials, but there may be exceptions there where you have past contamination with metals deposited from smelters or perhaps pesticides or other things even contaminating crustal materials.

We have provided in Chapter 9 extensive tables that indicate the risk values for mortality or morbidity effects per increment in short-term or long-term exposure concentrations of different PM indicators, PM<sub>1</sub> 0 , PM<sub>2.5</sub>, and

so forth, and for U.S. and Canadian studies. We have tried to highlight and separate out or identify separately those studies that were treated in the '96 document versus the newly-published studies assessed in this document that add to that base. Obviously, we are going to have to modify those risk estimates in light of whatever comes out of the future revisions or reanalyses of the studies that we are talking about.

Again, we think CASAC comment on Chapter 9 would be very helpful, still, for us to have comments on the overall organization of the chapter, how it is structured and, you know, addressing questions related to the flow of the risk assessment paradigm, comments on the level or depth of treatment of various topics and the extent of integration and specific ways to improve the integration across subject areas, soundness of bottom-line conclusions, especially regarding quantitative estimates of PM health risks, and the likely mechanisms underlying such effects and likely susceptible populations.

That is a quick run-through of some of the key revisions and issues or whatever that may help to focus some of our further discussions today. I think the next thing on the agenda, then, is to perhaps go ahead and move into having some presentations on the statistical issues that, I think, will be very important for all of us. I appreciate your attention and look forward to what we learn from these people.

**DR. HOPKE:** Any quick clarification questions that anyone wants to address? Roger?

**DR. MCCLELLAN:** Back earlier, you mentioned adding just about six months of data in terms of the concentration. What time period of data are you going to be able to go up to in terms of that additional data, as you reference on page 3, I guess, the third paragraph down?

**DR. GRANT:** Joe, what do we anticipate, possibly, on additional data that we might be able to add in on the PM speciation?

**MR. FLAAK:** Joe, could you come up to the microphone and also identify yourself?

**DR. GRANT:** This is Joe Pinto from my NCEA staff.

**MR. PINTO:** I would anticipate on the order of two to three months.

**DR. MCCLELLAN:** So, it will go through what time period, what calendar time are we updating for?

**MR. PINTO:** I prefer to defer that to somebody from OAQPS.

**DR. GRANT:** Roger, I am not sure we can answer your question in great specificity right now. Obviously, there is new data coming in from these various sites, you know, around the country as they have started up and begun collecting the data. I think we have in there, if I remember, Joe, '99 through 2000. So...

**MR. PINTO:** That is correct.

**DR. GRANT:** What we are hoping is that we are going to have additional data that would be available, quality assurance data and other things, that would take us three, maybe six months into 2001, you know, as possibly being able to add in there. We may be able to go further than that, but we' ll just have to see, you know, what is available.

**DR. MCCLELLAN:** I would just like to urge you to do what you can to shorten up that time in terms of the quality assurance and be able to include as close to contemporary data as you can. It would be great if...it is going to depend on when this full document is finalized, obviously.

**DR. GRANT:** Sure.

**DR. HOPKE:** Anybody else?

**MR. POIROT:** I had a general question that is kind of on the same topic, but I...a little bit difficult in some of the review aspects to not try to put things into a bin of can we look at this a little bit more in the future, but now, you have opened up that door, and I am wondering both from Bill and from Les, you know, how much do we really want to kind of direct you to things to look toward in the near future and how much do we want to really kind of say we generally understand things pretty well without needing to do that new work.



**DR. GRANT:** Right, and that is a very good point, one that I had in mind to touch on. Obviously, we are going to have to come to a point of wrapping up what period of time, some cutoff point for, you know, discussing new findings and so on in this document, and, you know, the statutes are such that whatever we don' t get this go-around can go to the next document unless there is something extraordinarily new or important or whatever.

We have in mind to, basically, cut this off as of the end of April which is, basically, what is covered in here. The most thorough coverage is up through December of last year that we were pretty good, pretty confident we captured most all the information there, and then we were able to capture some things coming out early in 2002 up through the point where we put the document out.

Obviously, you can' t have everything of the very last minute publication in a document of this size. So, we would like to go back and have a look back through and see if we can identify any really important new things published early in the year here up through, say, April to incorporate and then, you know, we have in mind to use that as a cutoff date for what goes in the next draft.

I think the air quality analyses, the AIRS data base, to the extent that is, as always, a little behind as far as things coming out, and, obviously, that data we are going to be talking about or whatever has to do with the de facto...or fresh, de novo analyses of these, you know, to the

extent that we can bring in more time, you know, for those or additional data, we think that would be useful to help extend, perhaps, what we have there in the way of being able to talk about, for example, PM<sub>10-2.5</sub> patterns.

**DR. HOPKE:** But again, you know, we need to bring things reasonably to a close, and, you know, again, if people see something which you feel is really substantially...begins to substantially change our viewpoint or does something major, then let's make sure to bring it to Les' attention, but, you know, we have got a lot of information already such that one or two more papers that say much of the same thing really isn't changing our basic understanding. So, you know, if there are things that look like they may be changing our basic understanding, really finally nailing something down or something like that, then let's get those few in, but otherwise, we really would be better off closing off around April.

And, you know, again, they are going to have to put in a lot of effort, as you will hear, with regards to chapters 8 and 9 to get those back together again so that, you know, at this point, I think we would be better off focusing their limited resources in that area.

Yes, Ron?

**MR. WHITE:** I am confused now, because you are talking about an April, 2002 cutoff for published papers which is essentially what is in the document right

now. **DR. HOPKE:** Well, it is really more December 30 with a few into early March.

**MR. WHITE:** But, essentially, we are saying April, 2002 is essentially unless there is something really earthshaking that comes along. On the other hand, it is very clear, Les, that you are going to have to incorporate some of the analyses, some of the key studies that, because of the whole GAM model problem, and it sounds to me, on the face of it, as though those are conflicting statements in that those reanalyses certainly will be post-April, 2002, and...

**DR. HOPKE:** Absolutely.

**MR. WHITE:** ...at some point, your head is going to have to cut off what reanalyses you are going to include to a stop point.

**DR. HOPKE:** Right.

**MR. WHITE:** And I would like to get a sense...

**DR. HOPKE:** And, to some extent, part of that is going to be a judgment on which are the critical studies for which we really need the reanalyses and which are yet another of, you know, maybe more limited value.

**MR. WHITE:** Right, and, you know, this may be getting into the discussion.

**DR. HOPKE:** Right, and that is really where we want to go tomorrow morning.

**DR. GRANT:** Yeah, I think tomorrow morning, after we hear...

**MR. WHITE:** But I think in terms of if you could lay out that in terms of that discussion, the time frame that you have got in mind and whether it...the identification of those studies and so on, I think that would be helpful to the discussion.

**DR. GRANT:** Yes.

**DR. HOPKE:** Right. Okay.

**DR. MCCLELLAN:** There was reference made to one other document, i.e., the Diesel Document that many of us spent a significant portion of our life span on. What is the current schedule for publication and release of that document?

**DR. GRANT:** I understand within about the past month, a final sign-off of the document has occurred as final and should be forthcoming very imminently.

**DR. HOPKE:** Yeah, I' ve had some recent correspondence with Dr. Feiland that he indicated that Dr. Gillman had now signed off on it so that it should be out here available shortly.

Okay, let's move on, then. Our first presentation is going to be S-Plus 101 by Lucas Neas who is going to help us statistically challenged folks to understand where we are going here.

**DR. NEAS:** My post-doc is

passing...post-doctoral fellow, Dr. Spenson, is passing out some splines. Spline is...I don't mean to indicate that this is not a serious matter. This is a very serious issue, and we have spent a great deal of time over the past eight weeks understanding this, and this is a very important issue to the Agency.

I'd like to express my gratitude to Dr. Burnett and Dr. Dominici who spent a great deal of time identifying this issue, and I'd also like to thank all of my extramural colleagues who have spent a great deal of time reanalyzing some of their data sets.

The purpose of my talk is not to foreshadow the later presentations. It is not to comment on the later presentations. I want to present a neutral introduction to some rather difficult statistical issues so that these become tractable to everyone here. So...

My name is Lucas Neas. I am with the Epidemiology and Biomarkers Branch of the Human Studies Division of the National Health and Environmental Effects Research Laboratory.

Let me introduce, although you are probably all well aware of this nemesis, the time-series study. The time-series study, very simply, is an epidemiologic study of the day-to-day variation of mortality or some other adverse health effect in an unenumerated open cohort where we compare that with the day-to-day variation in some pollutant of interest, generally measured at a central site, while

adjusting for the important other potential determinants of the adverse health effect, such as time and weather.

Many of these time-series studies have, quote, reported associations between adverse health effects and airborne particulate matter. Well, what are the alternative explanations?

In the epidemiologic study, we have a series of about five other explanations besides oh, it's the truth, the first of which is some selection bias. Very early on, this was a major concern that somehow the cities or time periods have been selected in such a way that it was true for these but not generally true.

Information bias is also a possibility. Somehow, the reporting of the health effects would depend in some unknown but non-causal way with the pollutant of interest.

What is more of an issue after these were dealt with is the issue of confounding. Does the selection and monitoring of covariance really effectively consider these other potential determinants of the outcome of interest, or have we somehow mis-modeled the data to produce an effect where none is present?

Finally, there is just chance. Maybe in the limited number of studies that we have conducted, just by chance, because of sampling variability, we are getting the odd study that shows an effect, and if we really could analyze everything, then we would understand that there really is no association. We are really getting just a few

studies by sampling variability that show an association.

Finally, there is the whole area of misclassification of either the outcome measure, the exposure, or the covariance. I am not really going to treat with that.

Today, it is mostly going to be confounding and the issue of chance that we are going to be grappling with.

In this, we generally, we have basically a multiplicative rate model. We believe that, in some category, the number of deaths divided by the person-time at risk, the rate of the event, is equal to some baseline rate times some rate ratios, that we have a multiplicative model for the outcome of interest which is the rate of, usually, mortality.

It is hard to model a multiplicative process, so we want to convert this to a linear model. So, by doing that, we take the logs, we move the person-time from the left-hand side of the model to the right-hand side of the model, but it is still the log number of events is equal to some offset term plus an intercept plus some regression parameters that can be interpreted as the log rate ratios.

This handles linear models very well. The linear effect of each covariant  $X$  on the dependent variable, the log of deaths, is related by some constant  $\beta$ , and this is the one we usually use for any sort of pollutant term of interest that we get,  $\times \text{F g/m}^3$  change in air pollution is related to proportional change or so much percent change in mortality.

Now, while this captures linear terms well, it doesn't capture nonlinear effects very well. So, generally, we have gone to generalized additive models where some nonparametric smooth function can be added to the linear effects of the other covariance. It is still additive. We assume the effects of weather could be added to the effects of temperature, could be added to the effects of time, could be added to the effects of air pollution.

And in most of our models, the person-time offset term is generally omitted from the analysis, because we are dealing with a large population where the amount of person-time at risk is really independent of the exposure of interest.

The reason that we need to treat with nonlinear models is there is a lot of nonlinearity out there in the world. Particulate matter has not been declining linearly over time. We had...this is data from Philadelphia from '73 through '80. We had an initial sudden drop in PM followed by more modest reductions in PM. Similarly for mortality in Philadelphia, we see this trend where there is a drop in mortality, probably due to a decline in the population of Philadelphia.

We also have nonlinear effects for season. We know that particulate matter in Philadelphia tends to peak in the winter. We also know that mortality tends to peak in the winter.

We also have nonlinear relationships with



temperature where particulate matter seems to be high during the cold or hot periods, probably due to its relationship with air pressure, and we have an effect of temperature with increasing mortality, in the case of Philadelphia, as the temperature drops.

We have to effectively deal with this nonlinearity in these crucial potential determinants in the outcome of interest before we can get correct effect estimates for the parameter of interest which is the air pollution. Linear models alone are not sufficient.

There are many different ways of dealing with nonlinearity. One of the ways of doing it is with a fully parametric model. That means we have some linear parameters, and although each piece of it is linear, in toto, they can trace a complex curve.

One of the ways of doing it are polynomials, just increasing high-order degree of polynomials, B-splines which I will explain more later, and N-splines or natural splines.

On the right-hand side, you will notice some abbreviations for these. The later presenters may just use ns and assume that you know that they are talking about a natural spline.

And there are nonparametric smoothers such as the locally weighted regression smoother, LOESS, and various smoothing splines. Smoothing splines are not often used in the analysis, and I won't treat much with them.

Splines. You are all holding the original spline.

A spline was a small sliver of wood that was used to trace a complex curve in an architectural engineering drawing. They would place this down on the drafting paper, pin in to the paper at various points. In between, they would flex the spline of wood in order to get the curve.

The pins, noted in green, are very important and are analogous to the statistical idea of a knot point. The choice of where you put the pins is very crucial to your ability to create the curve.

The difference between a B-spline and a natural spline is that a natural spline, like this piece of wood, at the end at the boundary point where the last pin is, beyond that boundary point, the piece of wood takes a straight line. That is the natural property. So, a natural spline model reproduces this natural piece of wood.

A B-spline is a piece-wise cubic regression spline, and they retain their curvature at the boundary pins so that if there was curving just up to that boundary point, the expectation of it would be it would still continue curving beyond that boundary point.

LOESS smoothing is done entirely differently. LOESS smoothing is a series of local regressions done at each data point. So, at this illustrative data point here, I am going to construct a regression smooth that will show the relationship between the independent variable and the dependent variable as a weighted regression using tri-cubic weight functions which give fairly high weights to data

points close to the data point of interest and relatively low, dropping to zero, at all the points that are at the boundary of the span and beyond.

This span is about equal to about 20 percent of the data. This is related to some sort of motorcycle as a data set, so we think that the relationship between this independent variable and the dependent variable at this point would be a point here. We repeat that for all possible points here and trace out the line produced by the stars, and that is a nonparametric approach. Other than specifying the span over which I want to evaluate the local regressions, the investigator makes no further choice.

So, the big difference between LOESS smoothing and a spline model is that in a spline model, the investigator has to make certain choices about where to put the knot points, and this is crucial to the development of the natural spline curve.

Now, I would like to introduce some of the other features. I am going to introduce five different smoothing techniques for this motorcycle data. They all, for 10 degrees of freedom, do about as well explaining the motorcycle data, but the polynomial models...here, you have a tenth order polynomial. Polynomials have this weightiness that is very undesirable in a regression model. It is just not very smooth. We are trying to develop the smooth relationship between the independent variables and the dependent variables, and here we have a lot of just sort of

weightiness, and that is indicative of polynomial models.

A better model is the B-spline. A characteristic feature of the B-spline is here we have a defection by the B-spline model of a curvature here to the data, and it is continuing to curve up at the end of the...at the boundary point.

The default, if the investigator doesn't want to specify knot points, well, knot points have to be specified for a B-spline or an N-spline model. So, the default is that the regression software just drops knot points at quantiles of the distribution. So, at every...approximately at every decile of this distribution, the software would stick a knot point.

It is arbitrary, but it is still a parametric approach where the knot points are or could be specified by the investigator.

The natural spline has a rather advantageous feature. Instead of curving up at the end, it has to be linear at the boundary and beyond. So, it tends to settle down as it approaches the boundaries and becomes rather linear at the boundary conditions. That is the crucial difference between B-splines and natural splines.

A LOESS smoother, through the data, captures the data very well. There is no polynomialness. Like the natural spline, it does settle down at the end, but it is not forced to settle down at the end. It is not forced to be linear at the end.

And a smoothing spline does about as well as the LOESS.

So, what is this monster, S-plus, and what is its threat to humanity? S-plus is a great statistical package. Any sort of type that has been placed upon this package has been inappropriate. It is a great statistical package, although EPA does not endorse...nothing that I say will they...but, certainly, nothing EPA has said should act as a dis-endorsement of this wonderful statistical package.

It is an object-oriented computer language like C. It was supposed to be the statistical counterpart of C. It is widely used by statisticians for the development of new methods. Any new technique that comes out is going to be written first in S. And it was the only software capable of utilizing general additive models until the release of SAS version 8 in the year 2000, and it is the most commonly used software for epidemiologists to do time-series studies.

So, what does S-plus do? Well, you have got a model that consists of some linear variables, and then you have got some nonlinear smooth terms. The first thing that it does is split off the linear component of the nonlinear variables and then a nonlinear component. It is going to analyze these as a series of linear variables so that every variable in the model is treated as a linear variable, and then the nonlinear variables, the smooth variables, will be treated differently.

We are going to cover GAM model fitting later.

Since it is an object-oriented language, the gam function produces a gam object which then must be evaluated by another function. When evaluated by the gam summary function, we are provided with tests of nonlinearity, and you can also get a plot of the nonlinearity. Most of you have seen these sort of smooth plots.

You can also have the software create a generalized linear model summary. This will provide coefficients to the linear variables and linear coefficients for the nonlinear variables. These linear coefficients plus the nonlinear component are combined to create the nonlinear plot.

Okay, now for some more messy details. How does it actually get these estimates? Well, it gets these estimates by, first, initializing the model covariance in the progression software, it will form a linear dependent variable, and then will create weight functions that say, on the basis of the Poisson distribution, how important are each of my data points.

It will then do a weighted linear regression to take care of the linear variables, and after it has done the linear regression, it will create a set of partial residuals, the portion of the variables that are not explained by the smooth...by the linear model. To each partial residual, it will fit a smooth and then create a new set of partial residuals, and then, for the next smoother in the model, do another smooth.

So, this linear variable is fit. Then, in turn,

each one of the smooth variables is fit, and then you evaluate across all the smooth variables, whether there has been any change in smooths since they were last fit.

And you are going to check for convergence. If you have not met your convergence criteria, you go back up here to another linear regression and keep iterating through this until you have met the convergence criteria.

Now, you have met the convergence criteria. This is just a simple sum of squares comparison. You are going to pop back out and estimate the deviance in terms of the Poisson distribution. You are going to check for convergence, going to form a new dependent variable and a new set of regression relates that are based on this new fit here. Then, you are going to iterate through this.

So, you have two checks for convergence, one in the backfitting algorithm, the inner loop to this algorithm, and then, once here where you are evaluating the change in the deviance which is the outer loop.

S...the gam function was developed early in the 1990s and was designed for use on the personal computers of the time, and in order to make this run in a reasonable amount of time, the S-plus default was set at  $10^{-3}$ . So, the loops were stopped when there was less than a 0.1 percent change in either the backfitting criteria or in terms of the deviance. Once that condition was satisfied, it thought it had the maximum likelihood solution and popped out.

In a recent letter, Trevor Hastie says well, that

may have been a little hasty on my part, and perhaps, with the improvement in computing power, we should have gone to a stricter convergence criteria, and he suggests  $10^{-8}$  with an upper limit on 30 passes through the loop.

SAS, I've checked the SAS manual, and it has now software version 8, and it stops at  $10^{-8}$ . I couldn't see from the documentation that there was any sort of backstop to limit the maximum number of iterations if this convergence criteria was never met. The only thing I will say is that some of my other colleagues have come up with some much more stringent convergence criteria, and they will be presenting their work later.

So, there are two issues. One, in this model fitting, are we really meeting...getting to the maximum likelihood solution? And that is an issue that the later presenters will present.

Another issue is that applying this generalized linear model summary to get these coefficients for the linear variables, does this summary extract the correct coefficients? In particular, does it extract the correct standard error for the linear variables when you have nonlinear smoothers in the model?

So, are there any questions before the statisticians get up here?

(No response.)

**DR. NEAS:** Thank you for your attention.



**DR. HOPKE:** Thank you very much, Lucas.  
If we all sit patiently, Lucas will hand out the piece.

Before we get into the presentations, I think this would be a good time for us to take a quick 10-minute break. Now, anybody who knows me knows when I say 10 minutes, I really mean 10 minutes. So, we will reconvene here promptly 10 minutes from now.

(**WHEREUPON**, a brief recess was taken.)

**DR. HOPKE:** ...presentations from HEI to look at, particularly, the NMMAPS project, so I am going to turn it over to Dan Greenbaum.

**DR. GREENBAUM:** Thank you, Phil. I wanted to start, first, by thanking Lucas for a very nice presentation. Everybody was always wondering what a spline was in the first place, but that was very helpful.

In preparing for this presentation, I was reminded of the comment that Stan Laurel used to regularly make to Oliver Hardy, well, it's a fine mess you've gotten us into this time, Oli. And there is a lot of work that has gone on as a result of some very good investigation by some investigators, looking into their models and checking, and what we are going to try to do at the outset this morning is provide a context for discussing some revised results of that work and HEI's peer review of that.

If you haven't seen it, there is outside a package that is provided. It has got a cover memo from me. It

actually has the revised results from the investigators at Johns Hopkins and at Harvard as well as the initial critique provided by the HEI special review panel for NMMAPS, and what we are going to do this morning and, hopefully, audiovisual changes willing, within a good short period of time, first of all, ask Francesca Dominici of Johns Hopkins to present her results and what she has been finding as she has redone her analyses, then Joel Schwartz of Harvard who conducted the morbidity analyses in NMMAPS to present his results, and then Sverre Vidal who is the chair of the HEI review panel and a member of the HEI review committee will present some perspectives from the panel, having had a chance to review...iterate with the investigators and give some both technical comments and then comments on the implications of the results, and then briefly talk about our next steps.

I want to just briefly remind you what the NMMAPS project was and is, the National Morbidity, Morality, and Air Pollution Study, an attempt to address some of the issues that Lucas put up, some of the questions, about selection of cities and other things, systematic investigation of short-term changes in air pollution and health. This was designed, from the outset, primarily to allow the combination of individual city results, not just to look at every city, but to allow the ability, knowing that some cities are smaller, have less statistical power, others are stronger, allow a statistically valid combination of those.

It included mortality analyses done at Johns

Hopkins in the 90 largest cities in the United States that had particulate matter data and morbidity analyses in 14 U.S. cities that had daily PM<sub>10</sub> data. This was elderly hospitalization analyses.

The original results were reported by HEI in 1999 and 2000, and it appeared in a number of peer reviewed journals in the U.S. and in Europe as well, and there are continuing analyses in this data set of dose- response and heterogeneity which HEI is funding and would have been reviewing right now but for this other thing that came up, but we are continuing to support those as important projects.

What I am now going to do is ask Francesca Dominici to come up and talk about the work that she and her colleagues have been doing at Johns Hopkins.

**DR. DOMINICI:** Okay, we are good to start. What I would like to do is I will spend a few minutes on the generalized additive model, 1 of 2, and I would like to thank Dr. Neas for a really great presentation that will make my life a little bit easier, and then show you the update of the NMMAPS analysis.

So, the main findings...and this is a really key point I wanted to make. One...and that is the reason why we are here...there have been some recent reported issues and discovery with the use of these generalized additive models and their implementation with S-plus software. So, there are really two things. One is the generalized additive model as

a statistical method to analyze time-series data and the implementation of this method through the S-plus software and, particularly, the function gam, and then there is a second issue which Dick Burnett will talk about, and what he discovered was that by using generalized additive models, you can have, in this particular situation, an underestimation of the standard error.

So, you know, in parallel, when Rick and I found this problem with the generalized model, what we started doing was start to analyze the data by using a different model that is a fully parametric approach that Dr. Lucas Neas was talking about and so to compare the generalized additive model with the other methods.

I will give a little bit of introduction in terms of the single-city analyses, and one of the things...you know, one thing I want to make...one point I want to make clear is that all these issues we talk for the generalized additive model, lack of convergence, underestimation of standard error, really are issues because of the...we are really trying to estimate an estimated fact which is a very small in the presence of many confounders which are highly correlated. So, we are really doing a job which is difficult, and that is why we are having problems with the software which is available. If it were a type of rate which were ten times larger, all these problems, you know, would not matter.

And, finally, I will...and, you know, that is

already of high interest to really show you the NMMAPS reanalyses under two different dosimetry models, and, basically, what we found is there was no qualitative change in findings.

So, things we decided...and we are going to go quick, because now, after your gam 1-1 and time-series study, you should be very familiar with that, so, what we do here is, you know, the NMMAPS and time-series studies in general, our goal is to investigation association between day-to-day variation in air pollution and mortality, taking into account several confounding factors. And the most important part whether short- or long-term trends determine mortality and seasonality, and the goal is to estimate mortality rates associated with short-term exposure on the order of days.

So, one of the things to keep in mind differently from the cohort study, what we do here is really to estimate short-term effects. So, if there is a high level of air pollution yesterday and two days ago, is there going to be an elevated rate of mortality the day after or, you know, a few days later.

The statistical approach...and, you know, that is following up with the previous presentation, in a simplified way, is how we do that. So, how do we estimate association between day-to-day variation in mortality and day-to-day variation in air pollution taking into account confounding factors through what is called the generalized additive model? And this is the only equation that I have here. What

we do is we model the logarithm of expressed value of mortality on a particular day as a linear function of the air pollution. Beta is the parameter of interest, and then, this is our...you know, our nightmare in the last...you know, my nightmare, for sure, and these are the two smooth functions...we have many more than that, but these are the two smooth functions of time and temperature, and the number of degrees of freedom will measure the level of flexibility of the smooth function.

So, to give you an idea is that the confounding effect of time and temperature on mortality are nonlinear. So, to give you an example, we know that mortality tends to be larger at higher temperatures and at smaller temperatures. So, really, the relationship between mortality and temperature is in a U shape. That is why I need to include a smooth function here, a more flexible function, to take into account for that.

So, the idea here is to estimate this beta which is percentage increase in mortality per 10 units increase in PM<sub>10</sub>, taking into account time and temperature. Actually, really, the generalized additive model that most colleagues are using is much more complicated than that, but here, I just want to make the point that this is an air pollution effect, this beta/PM<sub>10</sub>. These are the nonlinear confounders.

Now, how do we estimate the smooth function? So, the generalized additive model has been the common choice.

The generalized additive model now is coming back just because we are being borne out, but, you know, some of the issues in the generalized additive model, but really, we don't know yet which of the two is the best one, you know, considering that, you know, we can find the best one.

So, really, the issue here is that there is a substantial uncertainty within a given city. So, for any given city, we try to estimate a small pollution effect relative to the potential confounding effect of season and weather. So, the idea how do we control for season trend and weather?

Well, there is not a best way to control for these factors. What we can do, we find a reasonable way, and then we can see how the results change under a ten alternative reasonable ways, and that is what we have been doing.

But, you see, the point is that how we control for these confounding factors really affects the estimates of GAM, the problem of the standard error, and so on.

So, to put things in context, what this picture shows, the black lines are the estimates of the smooth functions. So, the black components of this graph is the signal that comes from the confounding factors, and the red part of the graph is the signal that comes from the pollution effect. So, what statisticians try to do is they try to explain the total variation in mortality by confounding factors or by air pollution, and you can see that the signal of air pollution is orders of magnitude smaller than the

signal for confounders.

Not only that, but if you look carefully at...I am not sure you can see it, but there are some situations where the air pollution signal, the red line, and the black line, the confounders, are co-related. So, the higher the correlation, the higher they have to find, one with the other, to explain the variation in mortality, the longer it is going to take the software to converge. So, this is really why we are having these problems. Okay?

So, the red air pollution, we are trying to separate out that signal which is smaller than the confounding factors. So, what are, again, the statistical issues?

One, convergence. So, what we found was that the default parameter in the GAM were not adequate. We are not sure of the convergence of the algorithm. How I found that, well, I was doing an analysis where the degree of correlation between these two components were larger and larger, and, you know, if you think of these two components were like one on top of each other, the algorithm will probably never converge. The model will not be identifiable. Okay?

So, I was increasing the correlation between these two more and more and more, and I was getting the same answer, so then, I was getting worried about it. So, we went inside the software, and there was, you know, several default parameters. I mean, we changed 4, but there are 12 default parameters in the open source, and the GAM software is, you



know, five or six pages of codes. So, just to give you a sense, it is not something which is easy enough.

One thing we need to keep in mind, though, that they default parameters are not a direct input function of...you can change the default, but they are not an input, so it is not something that you are going to look at first thing that you will do.

So, anyway, but what we found was that the algorithm was not converging, and so, we performed a simulation study which I am not going to show today to see, okay, now, if the default which was set to -3 were not, you know, adequate, how we can go down to make sure that we are getting the right answer. I mean, you know, is it  $10^{-3}$ ,  $10^{-6}$  is fine,  $10^{-8}$  is fine? You know, it is actually  $10^{-15}$  is fine, so that is where, you know, we really lowered down from  $10^{-3}$  to  $10^{-15}$ , and the pooled estimate of the NMMAPS study moved from 0.41 percent to 0.27 percent, but then, I am going to hold off on this one.

This is the second issue which I am going to talk of very little, because that is what Rick Burnett will talk about. So, in parallel, Rick was finding out this problem on estimation of standard error with the GAM, and this is a core, by the way, independently, if you use the defaults of the convergence parameter, and it is actually, for us, whenever you have multi-cities studies, the problem of underestimation of the standard error is really a problem for

the single-city studies, but it is not a problem for...it is less of a problem for the multi-cities studies.

And then we said, well, now, we have the problem with the convergence, we have the problem with the standard error. Let's pick an alternative model which uses a different estimation approach, and let's redo everything with this alternative approach, and let's see what we find.

So, there is really not a true model alternative for the NMMAPS data. What we found is that the generalized additive model was giving less bias estimates than the generalized additive model.

So, this is really what is happening. So, these are the margin of percent of distribution of the effects. So, this cord seen on the positive line, the center of the cord is the pooled estimate, and how this cord was spread is how much certainty we have. Okay? So, the red cord was the original estimate which was a center of 0.41...I am sorry. The black cord...the black cord was the original estimate which was a center of 0.41. When we basically used the more stringent convergence parameter from  $10^{-3}$  to  $10^{-15}$ , what is happening is this cord of the black shifts to the red on the left. You see?

Now, on top, though, what I want to make...you see this 1-1, this is the probability that the relative rate is positive, and you can see it was 1 before, and it is 1 now.

Then, what we did on your right column, we said

okay, now, forget about GAM, and let's use a generalized linear model which is the fully parametric approach which does not use a backfitting algorithm. It uses just an iterative relative square. So, you can see that is a little shift which probably has to do with the less flexibility in the generalized linear model, but still, you know, we move from 0.27 to 0.21 percent probability that these effects are positive that this cord lies on the positive line. So, the effects are small, but it is there.

Well, just very quickly, the problem of underestimation of the standard error doesn't really affect multi-cities studies very much, and these are...I redid all the analyses by taking the city-specific standard error and then taking the same city-specific standard error multiplied by 2, take the city-specific standard error and divide by 2, and you see that the pooled estimate is very similar.

The reason is because the variance of this core is the within-variance plus the between-variance. So, if you underestimate the within-variance, that will be picked up by the between-variance, and so, the total variance will be the same. So, that is just to say that underestimation of standard error is really important issue for single-city studies and less of an issue for multi-city studies.

Finally, that is the really the point to say that the problem with the GAM, the bias in the GAM, it really depends on two factors. One is how large the beta is, the relative rate estimate, and how large is the correlation.

So, what this power is which is a token color that I made here. So, blue is bad. Blue means that GAM is biased. GAM is not estimating the right parameter. This is a simulation study. Okay?

So, you see when I move the relative rate estimate here from 1 to 0.05 percent, so how the relative rate estimate goes down, the bias tends to go up. Also, how the correlation between the confounders and the air pollution goes up, the bias goes up. So, smaller relative rate, larger correlation between the estimate, make the problem more difficult to solve and make the algorithm in the generalized additive model, you know, really tricky, and you can have a lack of convergence and bias estimate.

So, it is really...you know, this is why you will not...it is not that you are going to see the problem with GAM or you are going to see change all the time. It really will depend on what is the level of correlation between the covariants, how you adjust for confounding factors, and how small is the coefficient that you are going to submit.

So, now, getting to the NMMAPS analyses, what we did. So, what we did is we reanalyzed the entire NMMAPS data base by using the same exact what we were doing before and, you know, just using more strict convergence parameters, and then we reanalyzed it by using a different approach, the generalized additive model, with natural cubic spline which is what Dr. Neas was talking about which is an alternative method which is a fully parametric method and uses a

different estimation of precision.

So, the NMMAPS reanalyses have been done. There was approximately 30,000 models, and I can assure you it was a lot of fun. So, now, what this picture shows is the estimates from the generalized linear model versus...so, then, the new estimates...by new estimates, I mean the estimates from the generalized linear model, just because I want to be very conservative and I don't want to, you know, have to do with GAM for a while, please, and these are the old estimates, and you can see each dot is one of the 90 cities.

So, you can see, actually, I mean, these estimates are pretty good, lined up on the yellow line. It does show that, you know, under the old estimates, it was a little bit more larger volume than before, and is the problem with the standard error which Dr. Burnett is going to talk about, and you see this is the old standard error versus the new standard error, and it is the problem of underestimation. So, now, the new standard error is going to be larger, a little bit larger than before.

Now, in terms of heterogeneity...and it is also a very important issue...well, first of all, before, also, there was not very much heterogeneity. Now, there is even less heterogeneity, and the reason why there is even less is because we have even larger city-specific standard error. So, there is a little bit larger within-city uncertainty.

So, it is true that single-city estimates might

vary a lot, but there is so much uncertainty in that that, actually, the heterogeneity is not so substantial.

So, to show you the national maps, you see on top I have the national maps of air pollution effects. Now, the color scale, you know, moves from -4 to 4. That is the estimates of the relative rate of mortality. So, the yellow and red points are the positive estimates, and the blue are the small or negative estimates. These points are much larger as the more certainty we have above the estimates. Okay? So, keep in mind these are estimates. We don't know if these are true, but we can say how certain we are about the truth.

And you can see that the city-specific...and, you know, each of these estimates is obtained by just using the data for that city. Okay?

Now, you see that, you know, these city-specific estimates vary quite a lot. I mean, they go from -4 to 4 percent. However, there is a substantial uncertainty, and if you look, the cities with the blue dots are the smaller estimates. Why? Because the smaller estimates are also the ones with the larger uncertainty. That is why when we pool, we obtain a positive estimate.

Now, what is the boundary, the division estimates? How we do that, we do a spatial smoothing, and we take the data for each city and take into account for the spatial correlation between the neighboring city, and you see there is a substantial shrinkage, and now, these estimates vary

from 0.1 to 0.3 percent. So, these effects are small, and they are heavily shrunk toward the other mean. The reason why we do that is because there is so much within-city uncertainty that they shrinkage is substantial.

So, this one also shows the pooled effect for total mortality at lag1 under the three methods and under three different methods for pooling. How the 90 city-specific estimates were pooled, you know, also that depends on which...you can...what I am trying to say you can combine the information across the 90 cities many different ways by using different statistical methods and still derive that the red and the blue and the green is just to show you the pooled estimates by using just a fixed effect, run with the effect model, run with the effect model, and the more complicated machinery which is the Monte Carlo Markov chain method.

Now, these three are the one, the original one, you know, which are centered on 0.4. These are the GAM when you use the multi-cities convergence parameters. These are the generalize linear model. So, you see they go from 0.4, 0.7, 21, all positive.

These were also the reanalyses of two key results of the NMMAPS. These will show the maximum percent distribution for total mortality which is the red color, cardiovascular and respiratory mortality which is the black, and the blue one are, of course, is mortality. Again, this color seen on the positive line, all the effect for cardiovascular and respiratory disease is the largest,

exactly as before, positive that these effects are positive, at 1, almost 1, exactly as before. On the right panel, you have the pooled effect of PM<sub>10</sub> under...well, maybe...five different multi-pollutant models.

So, what we did is we estimated the city-specific effect by using PM<sub>10</sub> only and then PM<sub>10</sub> that would include ozone, and then PM<sub>10</sub> that would include ozone and NO<sub>2</sub>, PM<sub>10</sub> and ozone and SO<sub>2</sub>, PM<sub>10</sub> and ozone and zero. So, now you are getting a sense of why there were 30,000 models. And then, we pooled the 90 cities, and you see that, basically, the color is, you know, similar. All sit on the positive line. The percent estimates are larger than 0+1.

Finally, this is really, I think, one of the key features, because how I start...I observe, you know, this. However, we say the adjustment to the confounding factor is key. It is key, because it really explains lots of the signal. It is key, because it is really going to tell us how big the coefficients will be, how large the correlation will be, how much trouble we are going to get with the GAM, and also, besides that, we don't know the best way to adjust for...personally, I don't know...maybe somebody else does...but I don't know what is the best way to adjust for confounding factors.

So, what we did and what this picture shows is at the pooled estimates of the 90 cities under 10 alternative scenarios of adjusting for confounding factors, and the red



one is our estimates, our famous 0.21, and then the other one, what we did, you see that in the right one, there is a number 222 which means how we specified the number of degrees of freedom. In the center is your time and dew point. And that is 111 which is what...this means that I took less number of degrees of freedom than standard, so I control less strongly for confounding factors than before.

So, you see that if I control less for confounding factors, of course, I lead to a pollution more times to explain, and the effect will be larger. If I control more, this plot goes more. But the picture is that estimates can vary between 0.3 to 0.2, but there are very few there.

So, conclusion. So, then my study conclude that there are three important conclusions about air pollution and mortality. There is evidence of an association between acute exposure to particulate pollution and daily mortality. This association is strongest for respiratory and cardiovascular causes of death. This association can now be attributed to other pollutants, including NO<sub>2</sub>, CO, ozone, or to weather.

These findings are basically unchanged, are qualitatively unchanged. This color shift on the left a little bit when using GAM with six convergence criteria or Gillam.

That is it.

**DR. GREENBAUM:** If there are any quick fire power questions for people...go ahead. We are going to

try to have a little time afterwards, but...

**DR. MCCLELLAN:** You showed the slide in which you had the posterior distribution for the reanalyzed effects when you included the other pollutants, and early on, you noted that the air pollution signal was an order of magnitude smaller than the confounder effect. Can you tell us what the effect was on the individual other pollutants in this model as a result of your analysis? In other words, what is the effect of ozone, what is the effect of NO<sub>2</sub>, what is the effect of SO<sub>2</sub> or CO?

**DR. DOMINICI:** Well, first of all, let me say that the reanalyses of the main effects of the other pollutants are still ongoing. So, I am not ready to say how have been the change of the effect of ozone and the other pollutants on mortality from the older analysis to the new analysis, because we haven't finished that yet. But, previously, what we had...and that is in NMMAPS report number 2 where the main effect was ozone and then NO<sub>2</sub>, and there were like, you know, we haven't seen any major effect of other pollutants on mortality except an effect of ozone for some, but I cannot comment on the effect of the other pollutants on mortality data in the new analyses, because they are still ongoing.

**DR. MILLER:** You noted that your analysis for synthesis and other degrees of freedom that you used for time and weather and so forth. What kind of sensitivity

analysis did you do to come into a reasonable number? You mentioned in your example there are seven, you say, expected. What is your criteria for determining that? Because I would think that that eventually drives the analysis.

**DR. DOMINICI:** Right, so this is why, you know, I included time so that this was the key feature, and, you know, let me reiterate that. You are right. I mean, how you specify the number of degrees of freedom, how you adjust for confounding factors is important.

So, what we did, we said well, we specified seven, you know, numbers of degree of freedom, seven for here as a function of time and space and temperature, because based on exploratory analyses, they were, you know, adequate to control as much as we can, because we want to make sure that we are not cumulative to PM<sub>10</sub> effects for temperature. So, these were...was the first stop with something that we felt comfortable, and based on exploratory analyses, we were taking out from the analysis the confounding effect.

But, you know, there is really no better way to adjust for confounding or, at least, you know, our group, we didn't decide to take the fastest way, and so, what these plots show is that okay, what we do now is if this is what we think, and now, let's take ten alternative ways which are more or less reasonable which reflect more or less drastic control for confounding factors, and let's redo everything under ten alternative ways.

So, what these ten estimates show you is what is the result, the main results of the Mason argument and the results of ten additional errors, and what we see is that city 4 here is a little larger. Why? Because there, we are using a smaller number of degrees of freedom. So, it means that we are controlling for seasonal and temperature less drastically which, you know, for somebody else, might be the appropriate way to do that, but in this way, you see that this is a just a little bit less, you know. But still, I mean, I found this picture quite reassuring in terms of how the pooled effect is, you know, it is robust when adjusted for confounding factors.

**DR. MILLER:** Well, guess my point would be that it argues to go for a greater number of degrees of freedom for these particular variables, and when you have a 50 percent change in the estimate by it, I don't think that is...can be defensible for having on the left-hand side the smaller number. So, the way you would get at this would be a simulation analysis where you actually constructed distributions and not...you are having to work with real data, so I realize you can't extract all that, but it disappears.

**DR. GREENBAUM:** I am going to suggest that we are going to have further discussion on this point, and maybe rather than sort of taking all the time here, it is not that...it is not an unimportant point, but I think that we should move on to the next speaker.

We are going to do a little switch in technology here and turn on the overhead projector, and assuming we can get this...

I just wanted to make one comment before he starts which is to remind people...we'll talk about this a little bit later, but this is, in part, in response to Roger's question. This was an effort, obviously, to do a lot of these further analyses prior to this meeting. We have, from day one, on May 30th when we wrote our letter, indicated that our intention at HEI is to continue these analyses past this point with a goal...and that is in response to other pollutants, with a goal being a fuller report on the analyses and a full commentary from the HI review committee by this fall. So, that is...so, we have also been asking the same question, as have the investigators, and as Francesca indicated, 30,000 analyses is a lot of work to do, and they wanted to do it right this time. So...

Thanks.

**DR. SCHWARTZ:** Okay? So, I would like to continue discussing the reanalysis of the NMMAPS study and focus on...do I have no input?

**DR. GREENBAUM:** I thought he muted it. Did you mute it? Sorry.

**DR. SCHWARTZ:** So...and talk about morbidity analyses and maybe come back to some questions that were raised.

So, to review the issues, we have a lot of these studies that have been published. What, exactly, is the problem? The problem is, one, that the default convergence criteria in GAM happen to have been set too laxly. Well, that is easy to fix. Work. Not everything, but that is not a major problem.

The second problem is that the standard error estimated in individual cities were estimated using an approximation. You remember what Lucas said where it takes every variable you fit nonlinearly, and it goes off the linear part, and then it fits the nonlinear part separately? It turned out it only used that linear part in estimating the standard error, because it was a much computationally difficult job to do it right, and that was an approximation they made back in 1990.

So, the software really needs to be updated to get that right. So, that is a bit more of a problem.

So, the question is, what are the implications of these findings, one which Francesca identified and one that Rick Burnett identified, on the conclusions of the large, multi-city, time-series studies that have underlied the time period section of the Criteria Document?

Well, the convergence criteria issue. At this point, all of the large, multi-city studies have been reanalyzed with the stricter convergence criteria, so we know the answer to that, and you are seeing some of the results of that, and the answer is that the major findings still hold.

Not only is there still a positive association with particles, as you just saw, there was no change in the coefficient of particles when you control for ozone or ozone plus CO or ozone plus CO plus 2. It just didn't matter, and I'll show you the results now for other NMMAPS and later for some other studies, and you will see that that basically turns out to be the case.

What about the standard error issue? Well, Francesca made a very important point. The standard errors are underestimated in individual cities if you use GAM, but if you do a multi-city study and you combine all the results, the standard errors are really, basically, not affected in your overall estimate, and that is because, in the old analysis, we underestimated how valuable our individual city estimates were, but then what happened is it looks like gee, the variations from city 1 to city 3 in the estimate was bigger than you would have expected, given those tight standard errors.

And it is the overall variability of the coefficients across all the 90 cities that contributes to the standard error of the meta analysis, and now we just partitioned it differently. The within variance is a little bigger, and the between variance is a little smaller, but the standard errors don't really change in the meta analysis.

So, in terms of deciding whether we want to go with the GAM models or with the totally parametric models that don't have any of these problems, standard errors really shouldn't

be much of an issue. We should worry about, you know, which do we think are doing a better job of fitting the data, and there, as Francesca said, it is not clear yet. There are arguments on either side, and I'll show you some of those.

But, basically, the standard errors you saw before didn't change. The effects side method has changed a little, but the standard error from the NMMAPS study with the new natural spline model is identical to the one that was in the Criteria Document before.

Now, that is also true for the standard errors from the NMMAPS morbidity analysis, and we only had 14 cities in this analysis, and we had a tougher job, because all of our cities had daily data, so our within-the-city standard errors were smaller, but, nevertheless, if you look at the standard errors from the combined effect for hospitalizations for heart disease in the old GAM model and the GAM model with the new convergence criteria and in the model using natural spline, they really don't change, and that is true for COPD admissions, and it is true for pneumonia admissions as well.

So, the standard error issue is not really an issue in these multi-city studies. Okay?

You will also notice that the standard errors for COPD and pneumonia are a lot bigger than for heart disease, and that is because a lot more people per day get admitted for heart disease than for these respiratory conditions, so you have a lot more problems.

Now, Lucas did a great job explaining the splines.



I just want to add one little point, and that is what we do, we could fit a polynomial to control for temperature. We think temperature, you know, cold days are bad, and very hot days are bad. So, we think it might look something like this.

If I fit a polynomial, the problem with fitting a polynomial is polynomials have symmetry. You fit a parabola, the left-hand side and the right-hand side are identical. They are mirror images.

But the effects of cold and hot days don't have to be mirror images. Biology doesn't have to be so symmetric. So, that is why it is nice to have something that is a little less symmetric if you are seeing it, and what a spline does is we can put a knot point right over here. That is where we put the pin, and we can bend it differently on each side.

What we are really doing, all we are doing, is we are taking one polynomial here, and we are fitting a second polynomial over here, and we require that they meet up at the knot point, because it is sort of embarrassing if they don't. Right? So, we are subjected to that constraint, but all these natural splines, to demystify one more step, is we got one polynomial here and one polynomial here. If you want to control for season, you can have a separate polynomial for every three months of your, you know, time period. That is what those splines are.

The down side of spline...that is a lot better than simple polynomials, having these polynomial splines, but the

down side is that it can be sensitive, as Lucas said, to where you happen to drop the knot, and that is no different than, you know, people used to categorize their data to look at the dose- response, and everyone noted that, depending on how you group the data, you can make the curve look linear or nonlinear. So, everyone started saying well, let' s just use quintods. So, the computer picks the arbitrary points, not you, but that doesn't make the sensitivity go away, and it is also the issue here.

Whereas a smooth, you would take the little window like this, you take the average, take the average of all the points in that window, and then you just slide that window along, and you just keep on sliding and drawing a curve. So, there are no knots to fit, so that is the nice advantage of smoothers, and that is why they were invented, and that is why we like them.

Now, so what are the answers when we reanalyze the NMMAPS morbidity studies? So, for cardiovascular disease, the overall effect estimate, using GAM with the old convergence criteria, was a 1 percent change per 10  $\mu\text{g}$  of increase in  $\text{PM}_{10}$ . If we use the new GAM criteria, it is a 1 percent change for a 10  $\mu\text{g}$  increase in  $\text{PM}_{10}$ . If we use regression splines, these B-spline polynomials, it is a 1 percent change per 10  $\mu\text{g}$  in  $\text{PM}_{10}$ . For cardiovascular hospital admissions, it doesn't change.

For COPD, it changes as little, 1.9 percent down to

1.7 percent, and then with the splines, it is lower there. It is 1.33 there, but, again, highly statistically significant. The basic message is the same.

For pneumonia, there is very little change with the new convergence criteria using GAM, using smooth functions to control for season and weather, but when we use splines, then we do get a big drop in pneumonia. So, I want to come back to that.

But, first, let me mention that Lucas spoke about B-splines and natural splines, and natural splines are constrained to be linear at the end. So, it occurred to us that maybe that wasn't the right thing to do for weather where the extremes can really blow up. So, we refit all of our models using B-splines instead of natural splines and redid the meta analyses, and the effect estimates for PM don't really change. I don't know if the weather predictions change, but the effects estimates of PM don't change depending upon whether we use natural splines or B-splines, but I did prepare a table on that.

So, what will happen to the pneumonia? Well, it could be that we are more uncertain about that. We don't have that many hospital admissions for respiratory disease, and they could bounce around a bit. But it could be that it has something to do with the greater flexibility of smoothers versus splines.

This is a plot of the residuals of the pneumonia models in Chicago using a smooth to control for season, and

this is for a window of 200 days, and you look at the residuals, and they look beautiful. Right? Randomly sticking around at zero.

This...these are the residuals from the model when we use the natural spline with the same number of degrees of freedom at the LOESS curve, and, you know, they don't so nice. Now, this, undoubtedly, is due to the fact that the computer...the whole algorithm of dropping a knot every so many observations just happened to hit the wrong days but from that point relative to some, you know, epidemics or something like that, but that...and this is the only one of the 14 cities where we saw something like this, but it does illustrate the fact that smooth curve are a little bit more flexible, and you don't get those things.

So, it is one of the arguments for whether or not we prefer to stick to the GAM, the extended area, so it doesn't really matter in the meta analyses...or go with the natural splines. As Francesca said, I don't think the answer is perfectly clear, but that could illustrate one of the sides of the argument.

Fortunately, the conclusions of the NMMAPS study don't really depend on whether you go with the smooth functions or you decide to go with the natural splines.

It is not just the basic results that don't change. All the other things that we did with this data don't change. This wasn't in the original NMMAPS report. We looked at the sort of things that might explain heterogeneity, and we

looked at, you know, income and poverty and race and all these other things like John and Francesca did, and we didn't find anything. But later, we went back to our data, and in a subsequent paper with Collin Hanson we did find something. We found that the coefficient from the  $PM_{10}$  effect on hospital admissions for cardiovascular disease increased with the percent of  $PM_{10}$  emissions from highway vehicles, and this is now the result using the natural spline model, and you see the identical thing.

And these are the same bubble spots that Francesca spoke about. The bigger the bubble, the more confident we are that this number is really here as opposed to, you know, there or there. Okay, so, the small points have wide confidence intervals. That is because, to your eye, big things look big. Right? So, diffusing the other bars gave exactly the wrong message, and that is why these plots were invented.

So, we are still seeing the same patterns of heterogeneity in the second model...

**DR. MCCLELLAN:** Could you explain your scales there?

**DR. SCHWARTZ:** Okay, yeah. This is actually the regression focus. So, this 0.001, that is 1 percent increase in hospital admissions for  $10 \text{ } \mu\text{g} / \text{m}^3$  PM and that is 0.5 percent, and that is 1.5 percent. So, as you go from the low end here, you know, where you are at maybe

0.7 percent increase to the high end where you are maybe, you know, double that, getting close to...

**DR. MCCLELLAN:** What is the percentage down below?

**DR. SCHWARTZ:** This is the percent of PM<sub>10</sub> emissions from highway vehicles from the AIRS web site for the city in which we did the study.

**DR. HOPKE:** Based on the emissions inventory?

**DR. SCHWARTZ:** Based on the emissions inventory which is right...but, you know, I don't know that they are created right, because I don't know the terms, so I can't answer...so, we published this analysis showing that, basically, it looked like traffic particles were more toxic than average. That still seems to be true. So...

**DR. MCCLELLAN:** Does it also say that it looks like a large portion of the PM<sub>10</sub> effect is associated with that small portion of the PM that is coming from vehicles?

**DR. WHITE:** Yes, because the intercept is low.

**DR. SCHWARTZ:** Yeah, and the intercept isn't zero. The intercept is about, you know, half. So, if none of the emissions are coming from vehicles, you get about half the effect compared to if all of the emissions are

coming from vehicles. As a matter of fact, there are two rings.

**DR. MCCLELLAN:** We can still play all kinds of games with numbers, I guess.

**DR. SCHWARTZ:** Now, one of the comments that was made on the NMMAPS study was that not many of the individual cities were statistically significant, and if you looked at the new analysis versus the old analysis, a bunch of the key statistics moved around back and forth across 1.9. So, then, that is not really an appropriate way to talk about the NMMAPS study which was as hierarchical study designed from the start to produce a combined effect estimate, and, in particular, the NMMAPS mortality studies include large numbers of cities that only have on the order of 50 PM<sub>10</sub> measurements a year. They have one 6-day monitoring, and then, usually, the equipment breaks.

So, they don't have a lot of power. So, you would expect not a lot of the results would be significant, and the key statistics would bounce around because of that, and the power of the study is that there were 90 of them. Okay?

To illustrate that in our part of the analysis, we did restrict the cities with daily PM<sub>10</sub> data, so we have more power at each of our individual cities, and this is a plot of the old T statistics versus the new T statistics using the natural spline model which has none of the problems that were identified. Okay? And, again, the main design of the study

is to produce an overall quantitative summary.

So, whether or not one city is statistically significant or not really doesn't matter, but you see that with more data per city, we have more of the individual cities being significant. You also see that there are no cities that used to be significant but now...that were above 1.96 but now are below. There are no observations in that quadrant.

There is one city that used to be insignificant that we have significant, but, basically, things don't change. The insignificant cities stay insignificant; the significant cities stay significant, and the overall effect is overwhelmingly significant.

I want to say one final thing about the question of covariant control. Season is a very strange variant. Season...season doesn't do anything to you. Right? Season is a surrogate variant for things that happen differently in the winter than in the summer.

And what we need to do when we control for season as opposed to some real causal variable is we are using time as a surrogate to deal with the fact that, for some pollutants, the pollutants tend to be high in the winter and low in the summer. Other pollutants, it is vice versa. And there tend to be patterns of mortality or of hospital admissions that are the same way.

And the idea is that if we take out the seasons and focus on shorter-term fluctuations, we have removed that



potential confounder, but the issue is to leave the shorter-term fluctuations. That is to say we have some idea of the time scale in which we think it is appropriate to leave the fluctuations that may be or not be correlated with air pollution.

So, arbitrarily throwing numbers of degrees of freedom at seasons doesn't do that and can create some problems, and I just want to tell you an example which is unrelated to air pollution that illustrates that. You are all familiar with the heat wave that hit Chicago in 1995 causing a large number, hundreds of excess deaths in the summer.

So, this is a smooth curve of five years of Chicago mortality data surrounding that period with seven degrees of freedom per year which is the default that Francesca used to be conservative, and that point out on the end of her graph is when she used 14 per year. Is that correct?

**DR. DOMINICI:** Yes.

**DR. SCHWARTZ:** 14 degrees of freedoms are here. They were doubled, and at that point, the coefficients weren't changing anymore, by the way. And what you see is here is a winter peak, a summer trough, a winter peak, a summer trough.

And look at that. The seasonal fit is clearly explaining some of the effect that we know is due to a short-term environmental variable, namely, the heat wave. And, therefore, when we try to model the heat wave here, if

we use this many degrees of freedom, we would underestimate the effect of the heat wave. And if it can do that to temperature in heat waves, it can do that to air pollution.

So, we do need to be careful not to throw the baby out with the bath water and not to fit models that control for fluctuations on such a small time scale that they take out the effects of air pollution.

And I can show you another plot. This is from Houston with seven degrees of freedom per year, and you can see, again, there seems to have been a heat wave in Houston that we know explains it.

But, also, look at this. Here is a little blip that occurred in the spring. Is that the seasonal pattern of mortality? Can you see where it is at? Something happened at that particular time in that year, and if the latter, don't we want to let temperature or air pollution see if they can be the thing that explains that little blip happening at that time in that year?

And the other thing that you can see is this is the correlation between the number of deaths per day today and the number of deaths yesterday with a lag of 1, two days ago, three, four, five, six, seven, all the way out to thirty-five, a partial auto-correlation function. Okay?

Now, we think deaths are independent events. When I control for weather and season which produced positive correlations between how many people died today and how many people died yesterday, what I should be left with is small

correlation coefficients that fluctuate randomly around zero. But when I use seven degrees of freedom per year, I don't get that. They are all negative.

And any electrical engineer will tell you that if you take a digital filter and you use a two-speed cutoff, you will induce ringing at higher frequencies, and that is what we are seeing. We are inducing distortion.

So, there are other down sides to over fitting. So, we don't want to under fit, but we don't want to over fit.

So, I will end with that and take any questions that you may have.

**DR. GREENBAUM:** Thank you, Joel. Are there any rapid fire questions? Roger already asked one.

**DR. SCHWARTZ:** Oh, and I should mention we don't see any heterogeneity left in our data anyway, either, even though we can explain some of the heterogeneity with the percent of particles entrapped, so it is not that you shouldn't look to see if there are signs of that, but, in general, there doesn't seem to be heterogeneity left.

**DR. GREENBAUM:** Yeah?

**SPEAKER:** Joel, your Chicago example, as you know, the smoothers use far less approximate degrees of freedom to predict the same residual variance. Did you try to optimally model the Chicago data for natural splines? Because you said that you kept them both the same degrees of freedom, and I thought maybe if you added more knots, you may

explain that peak better.

**DR. SCHWARTZ:** Oh, for that pneumonia thing?

**SPEAKER:** Yes.

**DR. SCHWARTZ:** Yes, I could, I am sure, or, frankly, I think if I used the same number of degrees of freedom and move the knots...right...I could make that go away. I am sure that is true, and I wasn't putting that out there to say that natural splines are a terrible thing to do but just to show that natural splines can be less flexible and, occasionally, you can get into trouble with where the knots have landed. You could use more degrees of freedom, but that, as I tried to show here, if you use too many degrees of freedom for season, you start getting into other things.

So, you know, exactly what the best way is, I think, still remains to be played with, but I think the overlying message is that the answers don't really depend very much on that, that we get the same results whether we use the natural splines or the smooth functions, and they haven't changed the bottom the studies in the Criteria Document except that some of the parameter estimates of changed, and others haven't really changed to any significant degree, but some have, but the overall conclusion seems that way.

**DR. GREENBAUM:** Thank you.

**DR. NEAS:** Can I ask whoever muted the

computer to unmute it now? This person doesn't seem to be in the room.

In introducing Sverre Vedal, I wanted to just comment on two things. One is to be clear that the NMMAPS study, like every other HEI study, is subjected to independent peer review by our review committee who are not involved in the development of the study or the oversight of the study itself.

What the review panel has been looking at involves the revisiting of the analyses done in the original NMMAPS work that was published in 1999 and 2000. So, that includes everything that Francesca Dominici has presented this morning is in the document, and the first part of what Joel Schwartz presented this morning which includes his attempt to redo the numbers specifically for morbidity in the 14 cities.

They had not seen prior to this, nor would they normally be reviewing, the redo of a published paper, for example, the one relating to highway work or the discussions about degrees of freedom, although, as I think you are seeing, the issue of degrees of freedom is getting some fairly consistent treatment.

So, with that, I just wanted to make clear what the review panel has been looking at. All of the material they have looked at is in the packets that you have available to you, and with that, I will introduce Sverre Vedal.

**DR. VEDAL:** I am wearing a different hat now as chair of the HEI Review Committee for NMMAPS. I just

wanted to take you through just very briefly the review process and who is a part of this particular committee, then present some discussion of our look at the analysis and simulation issues revolving around the GAM issue, highlight and repeat a little bit what has been presented before regarding the findings of both the mortality and the morbidity reanalyses, our preliminary critique as a committee regarding both methods and findings, and then some quick next steps from the HEI that Dan Greenbaum has already touched on.

The HEI NMMAPS review panel is a special panel of the HEI Review Committee. This particular review panel is not the same review panel that reviewed NMMAPS I or II. This panel was constituted to specifically look at NMMAPS III which had to do with the dose- response relationship of PM and mortality.

Nevertheless, some of these members were involved in earlier reviews, and the panel consists of myself. I am just a doctor and epidemiologist, but we have three statisticians on this committee, Ben Armstrong from the London School of Hygiene, David Clayton from Cambridge in England, and Nancy Reid from Toronto. In addition, Edo Pellizzari, an exposure assessment expert from RTI; Dan Tosteson who is also a physician and academic, dean emeritus at Harvard Medical School; and Mary White, epidemiologist from CDC make up the committee.

What we do is...typically, what an HEI review committee does and what we have done with respect to this

issue does not differ substantially, first of all, we review the methods used and, in this case, both of the analysis and simulation aspects of addressing the GAM issue, review the revised NMMAPS results. We comment...and these are just preliminary comments that I will be presenting but are the result of the committee's deliberations on this, and then formulate some discussion about what the implications of these findings are and what further work needs to be done.

With respect to GAMs, I think we were impressed that the NMMAPS investigators were the ones to identify this, and it has been clear, really, from day one that the investigators of NMMAPS have been very conscientious in terms of the development and assessment of methods in looking at...in performing time-series analyses, and I think, as a result of that, they came upon this problem.

Many have been using S-plus in GAMs and very bright people have been using GAMs in S-plus and not stumbled upon the problem. We also have to commend them on the rapidity and the thoughtfulness with which they put out the reanalyses and communicated the results to the scientific and regulatory communities.

Secondly, it raises a big caution to us in the scientific community about using statistical packages and not just statistical packages specifically but any out-of-the-box methods, I think, in general that we need to be more thoughtful about using.

Specifically with respect to the GAM issue, you

have heard that the default conversion parameters that were used in S-plus were not appropriate for air pollution time-series studies. Clearly, that is the case in NMMAPS. We are getting some appreciation of how general this is with respect to other time-series studies, and we may not see the same degree of problem with other time-series studies as we saw in NMMAPS, although we well may.

We also heard discussion of the GAMs tending to underestimate standard errors. That is still an issue when looking at single-city studies which is really what the bulk of studies to date are based on. As was mentioned, the multi-city study issue may make this relatively moot.

At the moment, alternative approaches to smoothing or removing the temporal trends or trends of covariants would have to use alternative approaches at this point as opposed to GAMs. Now, when the dust settles, people may well have justifications for using GAMs again, but at this point, certainly, alternative approaches should be used, if for no other reason, as a sensitivity analysis.

It should be emphasized that the cohort studies are not affected by this issue, at least, the American Cancer Society Six-Cities study and the reanalyses. Now, one might have motivation for using GAM in these studies for addressing spatial smoothing, but at least up to that point, the GAM issue was not relevant for the cohort studies.

To reiterate the revised results, most of the...for the 88-city point estimates, most of those point estimates



changed, and it wasn't that they all sort of tended toward zero. There tended to be sort of a general shift to the left, and the result of that is you actually have a somewhat increased or larger number of cities whose effect estimates are negative or zero.

The estimates of mean effect, as you have heard, shifted downwards, and I won't go over the details, the specifics of those, but you saw that there was a substantial decrease in the estimate of effect when more stringent convergence criteria were used and then a somewhat smaller decrease when a natural cubic spline or a parametric approach was used.

And you have also heard that the changes in the estimates were greater in models with more degrees of freedom. Now, that is an interesting issue, and Dr. Schwartz brought up some issues that one needs to address there. Do you control for those confounders more and raise the specter of more bias, or do you control less and be less biased? That is a difficult issue to consider and is going to, obviously, get more play.

The point estimates for all of the lags are smaller than before, but they all remain positive. The largest effect is for lag 1, that is, when the effect, the mortality, is lagged one day behind the increase in pollution. The lag 0 and lag 2 effects which were also recorded were initially about half of the lag 1 effect. They are still about half which makes them very small for lag 0 and lag 1, and I'

return to that in a moment.

The effect estimates for cardiovascular and respiratory deaths continue to be larger than for total deaths and for non- cardiovascular or pulmonary deaths.

The revised effect estimates for  $PM_{10}$ , using the approach that the investigators used there, appeared to be unaffected by the copollutants.

And the broad regional trends in heterogeneity remain, and they are similar to those that were found in the original analysis.

With respect to the morbidity, again, estimates of mean effect for elderly hospitalization were reduced. The change was more apparent using the parametric approach, that is, the general linear model natural spline approach, than with more stringent criteria using GAM, as Dr. Schwartz showed, but the effect estimates for the improved GAM were less apparent than we saw in the mortality analyses, but, again in this analysis, fewer degrees of freedom were used than in the mortality analysis.

There was also some suggestion, certainly, in looking at the Chicago example and pneumonia, that the parametric approach may not fully control for temporal confounding.

What is our critique, our look, on all of this? First of all, and it has been touched on earlier, there is, at this point, no gold standard for how we deal with long-term temporal trends. All of the methods have strengths and

weaknesses. GAMs has a problem estimating the correct standard errors, and that remains. With the parametric approach, there are problems with trying to identify the correct specification of models that use natural splines. Where are the knots and how many?

This notion of there being no gold standard, I think, further motivates use of simulation studies, as has been commented on, and motivates presentation of sensitivity analyses in work like this, that is, how sensitive are your results to different approaches to controlling for temporal trends?

There is a lot more work that needs to be done. How do these models behave with the type of data sets that are being worked with? So, although this is a commonly seen statement, I think, in this area, it has particular relevance.

Well, let's get down to sort of the essence here. What has changed? I mean, is the world the same, or is it different after this GAMs issue?

We have heard some discussion of heterogeneity, and the panel still thinks this is a prominent issue. Clearly, the purpose of NMMAPS was not to look at individual city effects. You heard that the purpose of it is to try to generate an overall effect that is a relatively stable effect, but the estimates are there.

Even though the heterogeneity is not increased and, in fact, may be a little bit smaller, as you have heard,

because the uncertainty in the individual estimates of effect are greater, if you look at the individual city effects, you will see that there is a greater number of negative estimates and zero estimates. In fact, when I do the bean counting, it kind of changes from about 28 percent to 37 percent either negative or zero estimates of effect.

Now, you can take two approaches when you have got that. You can do as the investigators have done, estimate a mean effect. Others would have a lot of concerns about even considering to estimate a mean effect in the face of that sort of display of city-specific effects. That is, if you conclude that all this is sampling variability, random variation in the estimates, by all means, go ahead and try to estimate a mean effect. If this heterogeneity, however, means that we have got a lot of apples and oranges in these cities, the next step should not be estimating a mean effect. And I don't think the dust about this has settled yet, and the committee is still concerned about this issue of heterogeneity.

The effects, as you have heard, have always been small effects. They are smaller now, and in the face of the heterogeneity, they are smaller relative to the heterogeneity.

Another issue that comes up now in terms of what has changed is an appreciation about the robustness of these findings. I think, before, maybe there was less concern about how robust or little robust they were. I think, now,

we have got a heightened sensitivity to it.

One of the sort of pillars previously was that the findings were not sensitive to analytical methods. That is clearly not the case. They are sensitive to analytical methods, and that has to have some change in take somewhat on what you feel about the robustness of the findings.

The lag effects, which were not presented, show that the lag 0 and 2 effects now are as diminished proportionally as the lag 1 effects. Those effects now are very small. They are less statistically significant than they were previously, and that also has to make you think about the robustness. In a robust setting, you might argue that you would like to see lag 0 and lag 2 effects similar to seeing lag 1 effects.

In the defense of the NMMAPS study, I must say that they have taken a very conservative approach to lags, and that was by necessity, given what you have heard about the data that a lot of it is every six to eight day data. They have assumed now that the effect now is a lag 1 effect in every city, and it is clearly not that. It changes from city to city. So, if you are trying to estimate a mean effect for lag 1, you are weakening what you are going to conclude, because that is not a consistent finding across cities. But, nevertheless, the fact that we have got variations in lag now might make you question a little bit the robustness.

It is unavoidable now that there is a perception that time-series studies findings are less definitive. It is

unavoidable.

I find, personally, that this has been an instructive exercise and may, in fact, strengthen some of the findings given that. It also may put even more weight on the cohort study findings, the long-term effects studies, at this point, given, possibly, some concerns about the definitiveness now of the time-series data.

What has not changed? Well, the benefits of the NMMAPS systematic sampling scheme have not changed, and the committee has been very firm about this as being a strength of the NMMAPS data. In fact, there is really no other study that gives us that picture of effects across a large geographic region with this number of cities in an unselected way. And the unselected is really critical to this. It is really one of our few looks at this very tricky issue of publication bias, and for that, I think it is extremely valuable.

The two-stage approach to pooling results or generating a mean estimate is more valid than what one might have done previously. The Bayesian hierarchical approach to doing that is a relatively valid approach, and that is a definite positive that has not changed.

The mean effects persist, although they are smaller, and I already mentioned the issue about the copollutants which, again, appear not to have much impact.

The next steps, Dan Greenbaum touched on these briefly, and I will also just touch on them briefly. All of

the key analyses from NMMAPS are being revised. The issue will be how you define key here. A fuller HEI report is intended to be published in the fall of this year that includes a commentary from us, the review committee.

There are other HEI-supported time-series studies that have been done. The plan is for those also to be revised when the GAM issue is relevant.

And then, in terms of further work on the issue of the correct model and heterogeneity, there are collaborative efforts underway in the States, in Europe, and in Canada, and, in particular, this APHENA study, Air Pollution and Health, a European and North American Approach, is a joint HEI and European Community project.

I will stop there.

**DR. NEAS:** Any quick clarification questions of Sverre, then?

**DR. MCCLELLAN:** I just want to comment on almost your last sentence, Sverre. You said the correct model. It seems to me that one of the things that comes out of this is there is no correct model, and what we need to do is stress the analysis of the data sets with multiple models to help us understand what is going on.

**DR. VEDAL:** Yeah, I didn't mean to say that. I thought I had stressed the issue of there being no gold standard, and what I meant by that is, at this point, there is no correct model. There are better models than

others, and I think the use of multiple models is...the motivation for that is strong.

**MR. POIROT:** At one point, you indicated...I think you said the effects are smaller, and they are smaller relative to heterogeneity which seemed to be different from what I heard from Joel and...

**DR. VEDAL:** No, the heterogeneity...

**SPEAKER:** Repeat the question, please.

**MR. POIROT:** At one point, I thought he said that the effects were smaller, and they were smaller relative to the heterogeneity, and that struck me as being different from what both Joel and Francesca said.

**DR. VEDAL:** No, the effect estimates have halved. There may have been a small effect on the heterogeneity in terms of being slightly smaller but nowhere near in proportion to the decrease in the estimate of effect.

**DR. NEAS:** Paul?

**DR. LIOY:** Is the take-home message here that you are concerned about using the mean estimate for the 90 cities?

**DR. VEDAL:** Yes.

**DR. LIOY:** And that that is a serious new uncertainty in our review at this point in time?

**DR. VEDAL:** Well, I think it is a philosophy of approach, you know, and I tried to...that is



what I tried to make clear, is do you either look at these as all random variation in a single distribution, in which case, I don't think anyone would have too much problem in going after a mean estimate.

Now, given what we see in terms of the individual city estimates, does it make sense to argue that, that these are cities from a single distribution? And that question is open and should be discussed, but it certainly would give you a lot of hesitation in terms of doing that.

**DR. LIOY:** Maybe part of it has to do with the fact that we are dealing with  $PM_{10}$  but that  $PM_{2.5}$  would have local components that are much more variable than  $PM_{2.5}$  which seems to have much more regional contributions?

**DR. VEDAL:** Yeah, we don't know that. I would doubt it, but there may be some element of that. You may lose some of the noise and end up having fewer negative estimates and such. I don't think we know that.

**DR. LIOY:** Well, it is just a point. I am...I think I agree with you about the idea of running a little more cautiously on the 90-cities composite.

**DR. LIPPMANN:** I had a question but, first, a follow-up just for a second on what Paul said. If we look at the cohort studies, they clearly show that  $PM_{2.5}$  is more closely associated with annual mortality than the PM coarse particles, so there is something in what Paul was

raising.

But I want to raise a different issue, and perhaps it is really addressed to Dan Greenbaum. We have in the handout figures 6 and 7 which were not discussed by any of the speakers. They relate to the seven different regions where...and perhaps Sverre wants to address this. Is HEI prepared to say anything about the reliability of the comparison between the seven different geographic regions which is in between the 90 cities and the individual cities?

**DR. VEDAL:** I am not sure I quite understand what the last part...

**DR. LIPPMANN:** Is there...I mean, in the original report, it was clear that the Northeast was different, and can we still say that in the light of the reanalysis? Because I think that is an important point, and you did choose to put it in a handout and didn't discuss it.

**DR. VEDAL:** It is probably something that the investigators should discuss as well, but I didn't see that as a big change from the revision of the initial analysis.

**DR. LIPPMANN:** No, but because the magnitude of the estimates are different, it might change whether we look at those as real differences or just noise.

**DR. VEDAL:** Yeah, I am not prepared to sort of discuss that further. I think, proportionally, the differences remain regionally, and I don't think there was a

big change in that from the revised analysis to the original one. Clearly, what you are saying is right. The estimates of the effect are approximately halved in each region.

**DR. HOPKE:** Francesca would like to say something about that.

**DR. DOMINICI:** Yes, please. Yeah, actually, because I think that those accord with the comment about, you know, the pooling. Just because, you know, we were approximating the accuracy of pooling these 90- city estimates coming from the same distribution, so what we did, that is exactly what we did with the regional analysis as well. One thing we can do is that we can partition the United States in seven geographical regions, and assuming that these 90 cities now come from 7 different distributions, one for each region.

So, this was really to...actually, in this way, we are imposing a heterogeneity, because we are saying that the cities that are in this region come from one distribution, and the cities from the South come from a different distribution. So, although that is why you pull out from figures 6 and 7, you see that, actually, you know, still within each region, the estimates were, you know, quite similar, and there was not much heterogeneity even within regions. There was heterogeneity between regions, because that is how we classed them.

So, for that, you know, to pull out really...there was an additional statistical analysis to say can we really

pool, so we said well, you know, if we cannot pool, you know, within the United States, we can try to pool within each region, and the regional estimates, to address the question whether, as in the pooled estimates, you know, it is lower than before or whether that is what you have in figure 7, they are all positive, and exactly as before, the effect is larger in the Northeast region and the Southern California region.

**DR. MCCLELLAN:** But one has to be very, very cautious in looking at those regions, because that would...you know, I am not certain who originally drew that, but, quote, individuals who spent a significant portion in the State of Washington and the Northwest...I am offended to have Oakland viewed as in the Northwest. If you review...if you remove Oakland, California, one of your two positive values, from the Northwest, I would submit your Northwest is no longer positive.

**DR. DOMINICI:** You are right.

**DR. MCCLELLAN:** And there are a lot of other problems with that map. I am disappointed that the NMMAPS investigators so quickly picked up on the EPA's use of that map earlier which was useful at the time, but I don't think it should have taken on the weight that it did here.

**DR. DOMINICI:** Well, just a quick reply to that. First of all, the regional analysis was an additional sensitivity analysis, as I said. Secondly, the

map that I showed you was actually very different. In that map, there was an especial smoothing, and what we did, we brought a strain from neighboring cities based on how close the cities were. So, that was a much more refined analysis, saying that the air pollution effect in New York would be a weighted average of the air pollution effect of the neighboring city, and the neighboring cities were defined with respect to the geographical distance.

So, that, you know, I hope that would be, you know, more definitive.

**DR. MCCLELLAN:** You get an oversimplification, and it may be applicable to the East Coast but have little applicability to the West in terms of meteorological conditions having greater local effects.

**DR. NEAS:** Can I just comment on one thing? I mean, the original report used the map regions from the EPA. The goal here was to say well, given the change in the statistical techniques, let's redo the analyses exactly the same way we did them before, not saying now that that map is the be-all and end-all but only to say we better do it the same way that we did it the first time so we know exactly what the changes.

But then, beyond that, as I said at the outset and as Sverre indicated, there is continuing analysis on heterogeneity that is being funded in this data set in order to try and look at it which isn't constrained by this

particular map, so we shouldn't overemphasize that.

I don't know if you want to...how you want to...

**DR. HOPKE:** Let's let Fred in here.

**DR. MILLER:** I just wanted to comment that I think we ought to be guided on this issue of whether or not to consider mean and pooling going back to the compositional data over the different regions and the information that is available, and if you do that, PM<sub>10</sub> is not the same in all locations, so it would, to me, argue against pooling and doing a mean estimate, but I think the other information in the Criteria Document is there to address that kind of issue and let's not get swamped here by the initial reaction to the size of the defects and whether they should be supported.

**DR. HOPKE:** Right. Again, I mean, we want to focus primarily at this point on the methodological issues and, I mean, again, we don't have the bottom line new answers yet, so I think we have to be a little careful about trying to over-interpret what still may be intermediate results.

Warren?

**DR. WHITE:** I just have a clarification question. I thought I heard Francesca say that the statistical outcomes were considered in arriving at clustering in the map. Is that correct? The regionality was

chosen, in part, based on the statistics?

**DR. DOMINICI:** So, what we did was two different regional analyses. One regional analysis would just partition the United States, you know, into these seven regions based on what, it is my understanding, based on what the EPA did. So, what we did is instead of assuming that these 90 city-specific estimates were coming from one distribution, we were saying that we cluster these 90 city-specific estimates with respect to which region they were and doing, you know, similar and separate pooling within each region.

Now, because we relied as, you know, as was pointed out, there was really across regions, but this was just really to say there are really, you know, some striking regional differences in what we see.

Then, we had the more refined method, and what we did...and that is where I showed you before where there was two national maps. The map in the document, that was actually an analysis. What we did is we estimated the relative rate of mortality associated with particulate matter within each city by taking into account the relative rate of mortality of the neighboring city, and the neighboring city were defined based on the geographical distance.

Now, I fully agree that other systems like weather and climatic differences would potentially be suitable, but that is what we did so far.

**DR. LIOY:** Well, I am going to support

Fred's point totally. If you are going to do reanalysis on the basis of regions to try to come up with any kind of average and you want to do it for a region, I think it is very important that you start looking at issues of what the composition of the aerosol is, what they are exposed to. It could be very, very drastically different from different areas of the country. Even if you are in one area of the country, if you are dealing with a city versus a rural location, you are going to get drastically different aerosol composition, especially for PM<sub>10</sub>.

If we were doing PM<sub>2.5</sub>, I don't think I would be as much concerned, because I think we have a relatively constant aerosol, and I think it would be very useful to think...go back and have people really look at the map not geopolitically but as a...geochemically. All right?

**DR. HOPKE:** Okay. Let's give Joel a last word here, and then I think we...

**DR. SPEIZER:** Can I make a comment?

**DR. HOPKE:** Sure, Frank.

**DR. SPEIZER:** I want to pick up on what Paul just said a little bit. It strikes me that...I don't have the slides, obviously, that people have presented to see what all was presented, the details, but I think they all did quite well in terms of what the issues are. Now, I have the feeling that if we had Francesca have another year and Joel to have another year to work on this, we might end up at



exactly the same point we are at now and that, in fact, we as a group are going to have to say yes, there are issues, and we see what those issues are, now let's move forward.

And the question is, if HEI or others are moving forward with perhaps looking at speciation or yet other ways of looking at it, you know, I think we are going to be...unless we start to move forward rather than spend time on reanalysis, additional reanalysis, we are not going to make much progress, and I think that is the issue that Paul is raising, and I think we have got to go to a different level of analysis, not just try to continue to repeat what we have done.

**DR. HOPKE:** Okay, but I think that is going to be a good thing for us to discuss later this afternoon and tomorrow morning.

Joel?

**DR. SCHWARTZ:** Yeah, I just wanted to make a point about this, and that is that there are really two purposes of these studies. One is to ask the question, overall, is there evidence that particulate air pollution in the United States is associated with mortality or is associated with cardiovascular hospitalizations? And there, I think, pooling the data over the country is appropriate, because it is asking, overall, is something going on?

Now, that is very different from saying that well, that is the number that is going on, and it is the same everywhere. I think it is perfectly true that the particles

vary, and we think the effects may vary, and, in fact, I showed you an analysis that we did using some pretty crude data that certainly suggested even with the PM<sub>10</sub> data we have that we do see variations in the effect sizes depending on the particle composition.

But I think Frank is right that, you know, there is a limit to what we can do with this PM<sub>10</sub> data for which we don't have any speciation, and, you know, playing...you know, I played around with that, you know, emissions inventory, but, you know, really what we need to do is rely on the other studies in the Criteria Document that did have speciation data to help us draw conclusions about that and not try to, you know, torture this data beyond the point where it can be done.

So, for the overall conclusion of is something going on, it is appropriate to pool the data. For the conclusion of, you know, what is going on where, I think that is where it becomes inappropriate to use a pooled estimate. You have to separate out those two things.

And I would also like to point out to Sverre that while he may have still topical issues with the distribution of the data, it is also true that a kind, clear test of the hypothesis that the data is not drawn from one distribution is rejected. So, I think we have to, you know, put that philosophy against that test to a little bit, but I wouldn't say that this is the number that is the effect in the United

States. This is a tool for concluding that there is an effect and that the effect of different kinds of particles or particles in different regions have to be addressed in studies that really have the data that we can use.

**DR. HOPKE:** Yeah, I think we really have got to cut this off at this point. We will have had more discussion later this afternoon, so let's break now for lunch. Let's reconvene at 1:25, and we'll pick it up where we left off. We'll figure out how to work this out for the rest of the afternoon.

(**WHEREUPON**, a luncheon recess was taken.)

**DR. HOPKE:** If I can have your attention, please, we are reconvening with another short presentation by Dr. Schwartz, and then we will move on to Dr. Burnett.

**MR. FLAAK:** For members of the public who are presenting today and who have materials to hand out, I have collected some of those. For those of you who still have handout materials for me to present to the committee, I would like to receive those in the next hour or so. So, please give those to me when you get the chance.

Also, since we are probably going to run a little later than expected today, I expect we will terminate the meeting at about the time we would normally expect according to the agenda today, and we might to try to carry a couple of the public comments over to tomorrow. So, for those of you who are going to be here, public commenters who will be here

tomorrow, please let me know if you are willing to be available tomorrow morning to give your public comments rather than doing them later this afternoon, and we might juggle the schedule just a little bit.

Thank you.

**DR. SCHWARTZ:** Well, you all know the issue, so, as I said earlier, at this point, pretty much all of the multi-cities studies have been reanalyzed, so I want to present some of the other analyses besides the NMMAPS so you have an idea of what has changed and what hasn' t changed as a result of that.

So, the first think I want to present is the Six-Cities time-series analysis that you may recall. The Six-Cities time-series analysis is the only multi-city time-series analysis using PM<sub>2.5</sub> as the exposure method. So, you know, didn't claim a thing.

So, without further ado, since you know all of the stuff that is going on, what we have here is here is the percentages and daily deaths per 10 F g increase in PM<sub>2.5</sub>. These are the old results that were published in the Schwartz, Dockery, and Eades studies with the 1.5 percent increase.

Now, if we go with GAM and the new convergence criteria, then it is a 1.4 percent increase, so a relatively modest change. If we go with splines, with natural splines, so we don' t have any of the issue of standard errors

associated with them, but, again, as we said before, it is not clear which things have the problem of covariants control...then it goes down a tad, to 1.2 percent, but still, highly significant. The lower confidence bound is 0.8 percent change and not a tremendous difference from the original ones.

Again, if you look at the standard errors of the meta analysis, we see the phenomena we were talking about before. The standard error issue is an issue for the single-city studies, but the standard errors, you know, are really the same in the GAM model and in the natural spline model once you do a multi-city analysis.

So, the Six-City analysis, the basic results hold up for PM<sub>2.5</sub>. The numbers are slightly smaller.

**DR. LIPPMANN:** Excuse me, Joel.

**DR. SCHWARTZ:** Yes?

**DR. LIPPMANN:** I see why, you know, 2.5 is probably more important, but for comparison, did you look at it in terms of PM<sub>10</sub>?

**DR. SCHWARTZ:** Well, I think the relevant thing to do is to look at the coarse mass again, and...

**DR. LIPPMANN:** No, no, I know. I mean, in terms of our discussion with the NMMAPS reanalysis which was a PM<sub>10</sub> reanalysis, it would be interesting to...

**DR. SCHWARTZ:** Oh, to see if the PM<sub>10</sub>

numbers changed more. That is a good point. I haven't done that. We'll put that on the list, and I'll send it to the staff. Good point, and I haven't redone the coarse mass, but, you know, we'll get around to doing that.

The other Six-City analysis that I think is important is Francine Leahy's analysis in which we did the source apportionment models and looked at the association with particles from different source categories. So, here is the reanalysis of that data, and we haven't done that with natural splines yet. So, all I can show you is the GAM model with the new convergence criteria, but as you saw from the earlier ones, it is not going to change tremendously when we go to natural splines in this, but we haven't had the time to do that.

So, here are the old results, 10 F g of particles from traffic were associated with a 3.4 percent increase, 10 Fg in coal particles was associated with a 1.1 percent increase in daily deaths, and dirt wasn't bad for you.

In the new model, the results don't change very much. They go down a little bit with the new convergence criteria, but they are in the same ball park, and they remain statistically significant for the traffic and coal particles. They remain not significant for the dirt particles.

**DR. WHITE:** Dirt isn't just not bad for you; it is actually positively good for you. You know?

**DR. SCHWARTZ:** Well, yes, that is true,

but this confidence interval, you know...

**DR. WHITE:** Excludes zero. Right?

**DR. SCHWARTZ:** No, no, no.

**DR. WHITE:** Oh, that is not a...

**DR. SCHWARTZ:** Protective effect -2 percent...

**DR. WHITE:** Oh, oh, okay.

**DR. SCHWARTZ:** ...and the range goes up to +1 percent. So, it includes zero. It included zero before, and it includes zero now. And these are all in the model simultaneously. The regression model has PM terms for traffic particles and the mass from coal particles and the crustal particles and in the, you know, places where we had the residual, there was, you know, the residual component, but I didn't show the results for that here.

Now, in addition to NMMAAPS, there has been one other very large multi-city comprehensive study, and that is the APHENA study which has looked at 30 cities from all over Europe, north to south, east to west. Those 30 cities have about 50 million people living in them, and that study was published in Epidemiology. The cities were picked before any analyses were done. It is not every city in Europe, because that was not feasible, but it was representative of all of Europe and selected before any data was collected, and the selection is in the branch application, so you can see that.

So, that has also been reanalyzed, and this will be

published in a letter to the editor of Epidemiology. So, this is the APHENA analysis which is for PM<sub>10</sub>, and I am going to show you two different sets of reanalyses that were done. One is from the main APHENA paper, and that looked at the mean of pollution today and yesterday.

And I should mention one difference between most of the studies that have been done in air pollution and NMMAPS mortality is most of the other studies that had daily data haven't been in the U.S., or they looked at a small subset of U.S. cities. So, they were able to look at the effect of today's and yesterday's pollution simultaneously, and all of the studies that looked at multi-day effects find that the effects do persist for multiple days. So, another thing that adds to some of the noise in the NMMAPS results is, you know, the inability to include two-day averages which will give more stable answers. This doesn't have that.

And the other thing that was noted...I noted this in a paper I published a couple years ago comparing one-day analyses to two-day means...is since the effect persists for more than one day, the effect size is somewhat larger when you use a two-day average. So, the fact that this is somewhat larger than NMMAPS is not surprising and doesn't mean they are really telling you different things.

So, the originally published paper in Epidemiology gave an effect size estimate of 0.6%/10 F g of PM<sub>10</sub> with a highly significant association. These are from random



effects meta analysis.

The new convergence criteria doesn't change those results very much. It is still about the same, still highly significant. I think these results probably changed less, because they are two-day averages, although they probably also used two degrees of freedom.

There was a second paper that we published. Antonella Zanabetti fit a distributive lag model to look at what the cumulative effect of air pollution exposure over a month or so is. So, what she looked at was today's PM<sub>10</sub> and yesterday's and the day before all the way out to 40 days ago. And this is an unconstrained distributive lag model, that is to say, all 41 terms were in the regression simultaneously.

Now, of course, probably none of them were significant when you look at that. Right? The point is that the cumulative effect is obtained by adding up all of them, and this gives you...and that is where some of a bunch of these numbers is less and you can get a more stable effect that way, particularly if you do it across multiple cities.

So, the results that we published originally were that there was a 1.65 percent increase when you looked out over the last month or so, about 2.5 times as big as the short-term effect. When we redo that with the new convergence criteria, it goes down a little bit, and it is now 1.45, but it is still about more than twice as big as the short-term effects. So, again, this is remaining significant, remaining about 2.5 times the short-term effect.

The point size estimates went down by a little bit, but in this case, less than in the NMMAPS study.

There is one more APHENA study I want to show you, and this is Annalis Hecht's paper in the European H e a r t journal which I don't even know if you referenced in the Criteria Document, but you should. This is a multi-city study of hospital admissions for heart disease in seven large European cities, Paris and London with seven million people each, plus...plus, in addition to those cities, the entire country of the Netherlands. So, there are about 40 million odd people represented in this analysis.

And this one is interesting, because this is the only case I have seen where the number actually went up when you use the natural spline with the same number of degrees of freedom. So, I think the general pattern is certainly that they go down a bit, but, you know, occasionally, they go up, and it is interesting to see that, and, in fact, it went up a tad when we used the new convergence criteria and GAM. So, in general, things go down, but that is not always the case.

There is one other multi-city study I want to tell you about. I published a 10-city study shortly before the NMMAPS study looking at daily deaths at  $PM_{10}$  in 10 U.S. cities. Now, the strengths of this study were I had daily data in all of them, so I could use the mean of two days' exposure and get a little more stability in the effects estimate, and I also fit separate models for the warm season and the cold season in each city so that I could get separate

in case the relationships were different, and then I did the meta analysis of those 20 estimates.

So, that was in the original publication, and the results were there was a significant association. There wasn't much of a difference between the summer and winter estimates, and the overall effect size was about 0.6 percent per 10 Fg.

Now, the new results, rerunning those 20 models with the new GAM convergence criteria...I didn't have time to turn these, you know, into percent changes and confidence intervals, so you'll have to pardon me for the regression coefficient, but this is a 0.65 percent change for 10 F g. So, with the new convergence criteria, it hasn't changed. And that is the standard error from the meta analysis.

And then if I do natural splines, again, as seems to be the pattern, there is a bit of a reduction in the effect size but not a really large one. It is 0.5 percent per 10 F g, and you'll notice, again, that in the natural splines, we don't have any of the standard error issues compared to the GAM with the new convergence criteria. In the meta analysis, there is really no difference in the standard error.

One other point about that is since these are studies where I have more power in the individual study, again, when you take a look at the T statistics of the old model versus the T statistics for the new model with natural

splines, you see, again, the study where the individual city and season-specific estimates were not significant stayed that way. When they were significant, they stayed that way. We are not seeing big changes.

The reason we see more movement in the NMMAPS mortality study is that, you know, 50 observations per year problem.

The last point is that natural splines are not the only way to fit a fully parametric model to analyze events data. Colin Bateson and I have published a number of papers, Beady has published papers, Longley has published papers. A bunch of people have published papers on the use of conditional logistic regression to do case crossover analyses as a way of analyzing mortality data or hospitalization data, and the idea is for each person who is admitted to hospital today, there are days, not that long ago, where they weren't admitted to the hospital, where they didn't die. And you can take those as the control days and make matched sets and do case control studies on them.

And what Colin Bateson and I showed was that if you pick the control days fairly close to the event day...and we showed what that means...but not so close that you start getting into other correlation problems, then you get no bias, and you fully control for seasons by design. You don't have to argue about how many degrees of freedom to use, whether natural splines fit better than smooth functions. By design, if the control days are within a few weeks of the

event day, you have controlled for season, and we showed that with simulation studies.

We then...Lucas Neas and I published a paper reanalyzing the old Philadelphia mortality data and showed that you got about similar results from the Poisson regression when you used the case crossover analysis.

There is a paper that came out in January by Tom Bateson and I which had partly some more simulation studies, but then we analyzed air pollution and daily deaths in Chicago. I think you ought to put that in the Criteria Document since it doesn't have either the issues of standard errors or the issues of, you know, which way to control for season. It controls for season by design.

And we found associations of  $PM_{10}$  with daily deaths that were pretty similar to those, for example, in that Ten-Cities study that I showed you before using the mean of today's and yesterday's  $PM_{10}$ .

And then, as part of all these reanalyses that we have done, I went back and redid that for Pittsburgh, and the results for Pittsburgh are very similar with this case crossover methodology. I get about 0.5 percent change in mortality for a 10  $Fg$  change in  $PM_{10}$  using the case crossover methodology. So, there is a good simulation literature and a small but non-trivial literature of reports of using this case crossover methodology to get around all of these problems and the parametric functions to control for weather

elements.

So, I think I will stop, then, and take any questions you have.

**DR. HOPKE:** Okay, are there any...Sverre?

**DR. VEDAL:** Yeah, you mentioned a couple of reasons as to why, in general, in the time-series studies, in the multi-city time-series studies, there remained less of a difference with the GAM...changing the GAM constraints. How much of that, do you think, is a degree of freedom issue in your analyses, that is, using lesser degrees of freedom and how much may be something else compared to the NMMAPS, for example, where...

**DR. SCHWARTZ:** I think that...well, if you look at the sensitivity analysis that Francesca does, that gives us a very good idea. If she went to half the degrees of freedom...right?...then the coefficient goes from like 0.2 to 0.3. Right? So, there is about a 50 percent change there. So, I think that gives you a direct estimate.

So, these other studies are showing somewhat larger effect sizes than 0.3. So, I think the other part of the difference is due to the fact that they are using multi-day averages which tend to give you higher coefficients and more stable results. They don't bounce around as much. But I think using those two things, you can partition it out, I guess, simply because I am 50/50 between those two.

**DR. HOPKE:** Other quick questions?

Question in the back there? Rick?

**DR. BURNETT:** Joel, have you had an opportunity to try that new risk sampling/percent sampling method for case crossover studies that appeared...

**DR. SCHWARTZ:** No, we have a different method that we published in that 2002 paper to deal with that small, that subtle bias. That bias that Longley identified, that bias is really due to the traditional thing that gives you bias in a case control study which is that the controls are not samples from the risk set. Right? That is to say there are some days that can contribute controls, but they can't contribute cases, because they were, you know, before you started or you had some missing days or stuff like that.

What we showed was...I mean, you can do that more complicated sampling. What we showed is we do our sampling, and then you can estimate the amount of bias by taking the same model and fitting it to a series where there is one event per day. So, the true coefficient has to be zero. And then you get some number, and if you subtract that off, that gives you the end bias, definitely.

So, we showed that in that other paper, and that is the approach that we have used rather than the other sampling, but we showed that they give similar results. But you are right, there is a subtle bias in these case crossover analyses, and we just do simple selection of controls. You don't have to do this more complex selection of the control days or estimate the bias and subtract it.

**DR. MILLER:** In the last study, the case crossover, when you said you get your control for season by design, you said you proved this by simulation. Do you mean sensitivity analyses, or you actually created an artificial distribution and did a true simulation? Because I would submit that on the cusp of any season, I would like to know how you prove that you could wholly control for season.

**DR. SCHWARTZ:** We took a bunch of different seasonal categories using sign and Sordo functions and sign and soil functions modified by various other terms and multiplied them, and then we took smooth functions of hospital admissions for pneumonia which show lots and lots of seasonality, and we took those all as basic patterns. Then, we simulated random time data around that, and then we did 1000 simulations each of those and estimated the coefficients and did it for different sampling strategies so we could find the ones that, for all of those patterns, gave you unbiased estimates.

**DR. HOPKE:** Okay, thank you very much, Joel.

So, now, our next speaker will be Rick Burnett from Health Canada. Do we need to get the projector back on? Yes, here he comes to do that.

**DR. BURNETT:** So, are you just about GAMMED out? Like a GAM conference, you know?

Okay, well, I would like to talk more about GAMs



and their interesting properties. What I am going to try to do is try to illustrate some of these issues with the different models with a couple of data sets in Canada, and there is actually another data set, a Canadian data set, that has looked at fine particles and mortality that was published in 2000. So, there are more than one fine particle data set out there.

I would like to also acknowledge the contribution of a coworker, Dave Stieb, from Health Canada and Tim Ramsey from the McLaughlin Center for Population Health Risk Assessment.

Now, I would just like to go quickly over why we chose GAMs as our optimal choice of modeling and why so many people want to do that, and I will go quickly over most of the points that have already been made. They are, obviously, highly flexible. They can handle a lot of variety of missing value schemes, not only just six-day data, but you have situations where monitors might be down for a long time because of various reasons so that the pollution data is very irregular.

One of the issues that we are all concerned was that the results were investigator driven. People would say well, I know what the weather model is, or I know I am going to adjust this for seasonality or so on, and people were saying, well, what happens if you did something else? So, there was a lot of interest in developing a strategy that would let the data speak.

So, with these generalized additive models, there really wasn't a need to specify a possibly incorrect and, in fact, usually, in statistics, as in life, every time we make a decision, we make a mistake. That's my life, anyway.

So, what we would like to do is we would like to make as few decisions as possible and, therefore, as few mistakes.

One of the other problems is that in this area, when people first started out to analyze these time-series data, they were using 4A series which are a series of sine and cosine functions to model seasonality, and there were concerns there that sometimes, in epidemics or other issues, that there were non- cyclic, some of the mortality pattern was non- cyclic, and we have some data, for instance, which is very hard to drive that series to white noise using 4A series. So, basically, we wanted something else that was a little more flexible than 4A series parametric models.

And we wanted less investigator decision making. We didn't want to say it was a square term for weather, or it was a certain functional form, you know, for time and so on. So, we wanted to sort of let the data tell us what we were going to do.

4A series don't capture non-cyclic trends, and we also looked at natural splines and said well...actually, we were more interested in nonparametric splines than natural splines, and other speakers have talked about this.

We've got to make two decisions, and like I said,

every time you make a decision, you make a mistake. We have got to know how many knots we want to put in and where the placement of knots are.

And, in particular, we were concerned that because of these very irregular missing data patterns when you could have a year or two of missing air pollution data, how would these natural splines react to that? I am going to get, all of a sudden, a number of parameters in my model that I am not going to be able to estimate, and what is going to happen there?

So, basically, I personally said I don't want to go with natural splines. I like these smoothing techniques, and there is no preferred parametric weather model. We don't know...I am not a biometeorologist. I don't know how weather actually affects health, and I am really interested in it as a nuisance parameter. I want to control, in some sense, for weather, because I have a vague idea that weather can kill. I don't exactly know how that is done, but I want to have some confidence that, at least within any given analysis, I have made some attempt at controlling for weather, and I understand that it can be highly nonlinear, and I want a lot of flexibility in that modeling.

Now, here is an interesting chart. We published in the Journal of Air and Waste Management in the April issue a meta analysis of, you know, every time-series mortality study we could find, and we went back, and we have just written a letter, Petros, to the editor of the journal...I think you

must have it by now...on the effect of GAM in that paper, and this is part of this letter, this plot.

As you can see, this is the percentage of time-series mortality studies that use GAMs. So, '86 to '95, none. We couldn't find any. '96, 40 percent. '97, it goes up. '98, 2000, and the study was cut off in 2000 at about 85 percent of time-series mortality studies included GAMs.

Now, if we look at the percent increase in daily mortality attributable to a  $10 \text{ F g/m}^3$  increase in particulate mass, those studies that didn't use GAMs, we found ten of those studies, and the actual average risk, using this sort of meta analysis technique, was 0.42 percent. We found five studies that had fine particles. The average risk was 0.82 percent. So, these are studies without GAMs, and if we look at the studies that have employed GAMs as the main method of analysis,  $\text{PM}_{10}$  has gone up to 0.7 percent, and fine particles have gone up to 1.15 percent.

So, you can see that as the GAMs have crept into the literature and our thinking, it turned out, by happen chance, I guess, that the actual effect sizes went up. Now, this is not a scientific study, but it does indicate a sort of a concern about how these GAMs have actually affected our interpretation.

Well, I am going to try to explain as clearly as I can how these models are working and why we are getting a little bit of variation in the results.

What we are trying to do is, obviously, measure the relationship between air pollution and mortality, and a number of speakers have indicated that this relationship is relatively small. The problem is that time also affects mortality and that time has a very strong effect on mortality. Weather also affects mortality, and its weather effect is somewhere on the order magnitude of air pollution. However, time also affects air pollution, weather affects air pollution, and time affects weather.

So, what we have here is a very complex relationship of a number of variables that are highly correlated, and we want to tease out one little...one of these arrows in this very complex relationship. So, it is a very difficult job.

I am going to try to illustrate how these relationships interact with each other with a study from Toronto. We had daily non-accidental deaths from January 1st, 1980 to December 31st, 1994. So, in Toronto, that is about 40 deaths a day. We didn't have daily fine particle data, but in this example, I predicted fine particle data from six-day fine particle data off a dichot and used daily sulfate and COH measurements to predict a continuous fine particle series.

Daily average temperature was used to model the health effects of weather, and model complexity was based on driving the residuals to white noise, because I wanted to fit whatever kind of complexity of model I needed. So, the

residuals didn't show any pattern, and the weather variables had optimum predictive power while trying to avoid overfitting.

So, what we are trying to do is sort of a compromise. We want the simplest model we can find that explains the most about the data. So, we don't want to stick in a lot of extra knots or do a lot of extra smoothing if we don't think that we need it. So, here is the kind of criteria that we used.

And these results were first reported in the Journal of Air and Waste Management back in 1998.

Now, this is what really burns me up. Here is an example of a model where I have time in the data and fine particles. So, that is all I have. And I have a LOESS smoother, I have a spline smoother, and I have a natural spline. I have a very weak convergence criteria of  $10^{-3}$  ...that is the default criteria in S-plus...and I have a more stringent convergence criteria.

And you can see here that, with this simple model, there really is no difference between the effect estimate of fine particles on mortality. Convergence criteria doesn't matter, the type of smoothing doesn't matter, and in this case, the actual natural splines are a little bit higher, but also, the confidence intervals, the standard errors are the same. Okay? So, it is very interesting.

And when we first got into using and deciding

whether we would use smoothers, we did simulation studies. Unfortunately, we did simulation studies with this type of model, because I was concerned about the really nonlinear functions which were the time. And this is the kind of results we got, and we looked at 4A series, and we looked at all kinds of other things.

So, we were very interested in this problem, and we got this kind of results, and we said okay, well, it doesn't really matter what you do. These flexible modeling techniques are very good, and they have all these additional properties that are desirable, so let's use them.

Okay. Well, let's go on to slightly more complex models. Here is a model with fine particles. Okay? In here. And we have a LOESS smoother of time, and the optimal smoothing, amount of span that is used here is about 100 days, so that was the...needed to drive those residuals to white noise. They had no serial correlation. And we had a natural spline of temperature with four degrees of freedom which is the optimal number of degrees of freedom.

And I plotted out the LOESS of time...so that is the general smooth temporal pattern of mortality in Toronto...in two cases, one where I have the default convergence criteria and a case where I have the more stringent convergence criteria.

Now, you will notice, unlike many U.S. cities, Canadian cities are thriving, and more people are moving into them, and, therefore, more people are dying over time other

than Houston or whatever. So, we get a slight upward trend in the number of deaths, basically, proportional to numbers of population.

Now, I look at these two curves, and my first view is they are the same. And I don't know if you have read like the Saturday papers, you know, when they have that little cartoon, and they have got two cartoons, and they say there are eight differences, and I can only find seven and I can never find the eighth. You know?

Well, this is what...when you look at this, you can actually see slight differences. You know, for instance, in here, there is less of a peak than in here. There are some cases where there is a bigger dip in here than here, and so on. So, there are little slight differences there, and I certainly would have thought well, that certainly cannot be meaningful.

The other thing you can notice is that in terms of optimality, in terms of removing serial correlation in the data or having the best predictive model, there is a lot more that we want to take out of this data than just seasonality. So, you can see here that we get the typical peaks of mortality in the winter, but we also get little jagged points and so on and so forth like this. So, there is certain structure in the data, and this is the optimal model even with air pollution and the weather in the data...in the model.

So, if we want to look at day-to-day effects, there



is more than just seasonality that I think we should be taking out of the data.

Here is a case where we have...so, obviously, the convergence criteria does make a difference in these subtle estimates of temporal effects.

Here is a situation where we have three different models. We use LOESS to model time, we use spline smoothers, and we use natural splines, and weather has exactly the same modeling. Again, you can see that these three curves all look almost the same, but there are slight differences in those curves. Okay? And it is very important to understand that we are going to see those slight differences are going to be meaningful.

Now, this is an example of the relationship between mortality and temperature, and the top two panels are a model of fine particles, a LOESS of time, and a LOESS of temperature with 30 percent span which is the optimal span for that data set. And here we see that if we have a weak convergence criteria, the relationship with temperature and mortality is kind of a V shape. So, we have got this clear pattern that as I get colder, more and more people die. And the PM parameter here is 13.33 or about 1.33 percent per  $F \text{ g/m}^3$ . And a standard error of 2.555.

Now, if I go to the more strict convergence criteria...remember, there are these subtle differences in how I model time, but that translates into a difference in

the shape or the relationship between temperature and mortality. Okay? So, the effect of cold temperatures isn't as great. So, you have got to remember, now, that, you know, cold temperatures in Canada only occur in the wintertime, fortunately, because nobody would live there if they occurred in summer.

So, there is this strong correlation between temperature and mortality, and what we have here is a slightly less effect estimate of fine particles. And the standard errors now are about the same. So, we have used two LOESS smooths in our model.

If we go to a spline smoother, the same type of thing. The spline smoother has a greater tail for temperature here than in here with the convergence criteria, and as the convergence criteria gets better and better, we change the relationship between mortality and temperature, and we go down. As that relationship changes, we get less and less of a fine particle effect.

So, here is a case where, now, I have got these strict convergence criteria all the time. So, I don't worry. I have gotten rid of the problem of the convergence criteria, and now I am going from a LOESS function of time to a natural spline, and I am doing a LOESS function of temperature, so that stays the same. So, here, this relationship with temperature and mortality changes, and when I go to a natural spline in this case, I get a plateauing. I don't get any cold weather effect anymore. And now, I am going down from

10.97 to 6.61, so I am going down quite a bit, and I can start to see that the standard errors are starting to go up.

So, as I put in a parametric model for time, it starts to inflate the standard errors. If I have a smoothing spline versus the natural spline, the same thing. The smoothing spline is much closer, and the difference between the PM effects are much closer also, but, again, I get a much higher standard error.

And here is the natural spline with a natural spline for temperature and a natural spline for time, and this is a 4th degree polynomial, so this is temperature, temperature squared, temperature cubed, temperature to the 4th which happens to be the optimal polynomial to model temperature. And, again, you can see that these two curves look similar, and the effect estimates are similar.

And the last one of these plots is what I have done now is I have kept the model for temperature the same, and I have just varied the model for time, and you can see that whatever model I put in for time changes the relationship between temperature and mortality even though I am using a natural spline all the time for temperature. So, you can see how these subtle relationships are having an effect.

So, what is happening is the way that I model time-mortality relationship affects the way that I model the weather-mortality relationship which affects the way...my estimate of air pollution and mortality.

Unfortunately, this is model sensitive. How I do

that...okay...how I do that cascading modeling of very highly colinear variables ends up being somewhat model sensitive. And that is, you know, completely against the first slide I showed you where if I only had time and air pollution, I got exactly the same results. Okay? So, putting that weather variable in has really changed everything.

Now, I am going to just briefly go on and talk about an eight-city Canadian study of mortality and air pollution, and I present these results, because it is sort of like the type of study the NMMAPS analysis, and you guys are ten times the population and have ten times the number of cities almost. So, we only have eight cities where we have real fine particle data.

So, we have daily non-accidental deaths in eight cities from 1986 to 1996. It is an 11-year period. These are sort of the most populous cities in Canada.  $PM_{10}$  values were obtained on a six-day sampling schedule from dichotomous samplers. It turns out that, like NMMAPS, the previous day's exposure displayed the largest risk.

Now, we are dealing with fine particles, not  $PM_{10}$  here, and it turned out from our weather modeling that daily average temperature and the maximum change in barometric pressure within a day which we always find is actually the strongest predictor of mortality of any weather variable, so it usually represents a frontal activity...is in the model, and these results were first reported in Inhalation

Toxicology in 2000.

Okay, if we compare LOESS, the fine particle parameter estimates under a LOESS model with the weak convergence criteria and the strict one, you can see that being below the line here generally means that the LOESS with the weak convergence criteria gave us generally higher estimates than the more strict one.

The splines, actually, the smoothing splines were a little bit more variable, though they were sort of clearly lower except for sort of one point up here. And the LOESS with the strict convergence criteria and the natural splines, the LOESS still tended to be higher than the natural splines except for one city, and that city happened to be Calgary.

What happened in Calgary was that the actual air pollution data was quite consistent. There weren't really irregular patterns, but the way that these different functions modeled weather was quite different. So, the optimum model using a LOESS term and the optimal model in a spline term gave us quite different results, so you got a different air pollution effect. And the smoothing splines and the natural splines.

The standard error issue has been seen before. The natural splines tended to always give us a higher standard error, and as the standard errors got bigger, the effect of this bias became proportional.

Now, the summary results here are that here we have another  $10 \text{ F g/m}^3$  change in fine particles in the Canadian

cities, and if we do a meta analysis, a random effects summarization of the results...I apologize for this, Barry, wherever you are...the LOESS terms, going to the stricter convergence criteria tended to diminish the effect somewhat, the smoothing splines had a little bit bigger sensitivity to the convergence criteria, and the natural splines which really aren't sensitive to the convergence criteria at all, at least what we find, gave us a slightly lower result.

I mean, here is zero down here. Our general conclusions, obviously, aren't changing. We are getting smaller point estimates, but, obviously, they are highly significant.

Now, one of the intriguing things that we found in this analysis is that we could detect our estimates of heterogeneity of effect across the cities were positive when we had the LOESS functions, but when we went to either the smoothing spline or the natural spline, we removed all that heterogeneity. So, in this one example, it actually pushed the actual estimates closer together. So, that is even when we had these underestimates of standard errors in the smoothing spline, we still couldn't see any heterogeneity.

That is why we have to always reanalyze these data sets or I would actually prefer to do new and better studies is because, for instance, the standard errors on the LOESS of the pooled estimate were actually the same as the natural spline. Okay? Because there is extra heterogeneity here, and there is none here. So, you never know what you are

going to get. Things could change quite dramatically.

Now, Francesca mentioned...I talked about the GAM standard error issue, and we actually discovered this issue not looking at time-series data but looking at the American Cancer Society cohort data. I have a couple of slides at the end on that which I'll get into.

Models containing nonparametric smooth variables will underestimate the standard error of the air pollution effect if obtained from sort of general statistical packages such as S- plus or SAS. The amount of underestimation depends on the nonlinear correlation between the smooth variables and air pollution, and that nonlinear correlation in smoothing is called concurved.

So, what these packages do is they essentially assume that the nonlinear smooth variables are actually linear variables when they go to calculate the standard error, and they do that because it is easy to compute. So, they save computational effort in here. Obviously, in studies like this, that has a real impact on estimates of the standard error.

There is, in theory, an exact variance estimator, so we could do a correction, but it is extremely computer intensive, and there has been developed a less computationally intensive approximation which exists.

We have programmed that up, and I have sent that around to a number of people, and Lucas has actually been playing with this a little bit. In situations we have looked

at, we haven't seen any problems with it, but Lucas has actually found a couple of situations with kind of sparse data with a lot of smooths in it in which the actual approximation doesn't look very good. So, we are going to have to, if we are going to use this approximation at all as a correction factor, we are going to have to look into this very seriously.

In multi-city studies, as a number of people have indicated, there tends to be less effect on the standard error of the pooled estimate. That is certainly because we have two sources of error. We have uncertainty within a city and heterogeneity between cities, and we are just trading off those. As we inflate the uncertainty within a city, we are just trading that off with the heterogeneity.

The difficulty is that it can have a very large effect on our estimates of heterogeneity, and like I said, in our eight-city study, we thought we had some evidence of heterogeneity under one modeling approach. Under another one, we find no evidence of heterogeneity.

So, our conclusions there in these multi-city studies, you know, could really be different.

So, air pollution effects can be sensitive to model assumptions. The sensitivity turns out to be a function of the number of smooth variables in the model, the amount of nonlinearity and concurvity in the regression variables...for instance, if I put in a linear term for temperature and I apply all these different methods, I tend to get a small



amount of sensitivity but not very much. So, no temperature, all the results look almost the same. Linear term for temperature, they look a little bit different but are very close. As soon as I put in a quadratic term for temperature, boom, they all change.

So, it is this nonlinearity in the data that is driving this, and this has a major issue with modeling and weather effects.

One of the things that people started to do is they started to say well, weather is a potential confounder. It is another explanation for the air pollution effect. We' 1 1 put more and more weather variables in the model, we will make them highly nonlinear, we' ll stress the data, and it turned out that the more weather variables you put in, it did have some effect on the PM effect, but PM still survived...okay...using these smooths. And I suspect if we really, you know, hammered these data the parametric functions, the PM effect may be much more sensitive to that.

But that doesn't mean...like I can always throw a bunch of variables in. I can throw enough uncorrelated variables in, and I can explain all my data.

So, again, we have got to come back and revisit, I think, the idea of what are optimal weather models. How much modeling of weather do we really want to do? How much do we believe that going from 76 to 77E F is going to kill somebody which these which these current weather models suggest they do? So, I think this is going to be another area of major

research for us.

Sensitivity appears to be on the absolute scale, as has been mentioned. If the risks were 20 percent and we went down to 19 percent, nobody would care. Right? I would not have the pleasure of addressing you today. The fact is the attributable risks are around a few percent, so when they go from 2 percent to 1 percent, that makes a big deal.

And if you talk to people that work in smoothing in this area, you know, they'd say well, you know, you are really stressing the data. You are really stressing the data with these models. You are trying to find a very small risk with a lot of confounding variables, and you can't expect something for nothing.

So, I wouldn't be too hard on the developers of GAMs and the software and so on. It really is a difficult situation.

The properties of GAMs, although if you read the GAM literature, it turns out that most of the properties, whether algorithm, convergence, or knot, and a lot of other things all assume no concurvity, no correlation between the regression variables. So, here is an area that maybe is prime for statistical research is if we start to add in some concurvity, can we actually push through some interesting statistical results for these variables.

Natural splines should be given serious consideration. At the beginning of the talk, I said I didn't like them for a number of reasons, where we pick the knots,

the number of knots. Joel mentioned, and I concur, that sometimes these things can be very whippy, so they look like a roller coaster ride, especially at the ends of the data, and you really believe that that is a real, true biological effect or not, where the smoothing variables give you what we think of as, a priori, a much more reasonable dose-response model.

So, we are going to have to really look at that, and I think Mitch Klein is going to talk to us shortly. I think he has tortured some of these spline models and looked at that.

They do provide more flexible models in certain parametric approaches but maybe somewhat less flexible than GAM, so there may have to be some tradeoffs here. Possible limitation of handling irregular or missing values, I think that has to be a major concern, and we have to do a number of studies to look at that.

We need to stress natural spline in all reasonable situations. If they are going to become our major alternative or one of them, we don't want to have this kind of meeting again in a couple of years. You know, we have really got to take some patience. I mean, I would really suggest that we mount new studies as opposed to reanalyze old ones, that we take some time to look at this issue, that we don't rush in and say okay, we have got the answer now, and then, six months from now, we are going to have another answer.

And we need to develop an approach to evaluate

these methods in general, because usually, in statistical analysis, you develop a method that is optimal for certain reasons. You say this method is going to do this, and then I am going to see if that method does that.

And what we are asking here is we want one method or a handful of methods to do everything. We want to take any data situation, any amount of complexity, and we want to be assured that we are not too far off. And that is a lot to ask.

Then, for a specific data set, you go out and do a study, how do I know that the analysis that I am doing is not way off base? How am I going to figure that out? Is it going to be simulation? Is it going to be something else? And how do I actually simulate a true underlying model when I don't know what it is? Okay? So, there are really challenges here to this.

Now, I will just talk a minute about the American Cancer Society study. GAMs were not used in the original publication in 1995. The follow-up period was 1989 to 1989, and that was published by Arden Pope in '96. The GAMs were also not used in the HEI reanalysis, so Dan Greenbaum is quite happy about that, I think.

Again, it was '82 to '89 follow-up. We used several methods to account for spatial autocorrelation. In that follow-up period in the ACS study, there was very strong evidence of spatial autocorrelation. That really tells us something about what is going on in the data, but it also

confuses us about the statistical properties of our model.

In this case, what we did is a number of things. We had regional indicator variables. We used the dreaded seven regions analysis. We had region-specific analysis that we looked at so that we could try to reduce spatial autocorrelation. We did spatial filtering which is kind of like a spatial analogue to a 15-day moving average filter for time-series data, and we did things called spatial autoregressive models which is, again, an analogue in space of time-series models. But we didn't specifically use GAMs.

Now, I had the bright idea of saying after the reanalysis was over, to say well, if we want to remove serial correlation in time-series data by fitting smooths of time, then, if we want to remove spatial autocorrelation in spatial cohort data, why don't we model a surface and a mortality surface?

We did that, and we published a paper using spatial GAMs that we developed in Environmental Health Perspectives in 2001, and this method worked beautifully. It certainly did control for all the spatial autocorrelation that we wanted.

Okay, the JAMA paper of 2002, this was a longer follow-up period, 1992 to 1998, and what happened over the '90s in the ACS study is that, basically, the mountain of residual mortality in the rust belt below Lake Erie that occurred in the '80s was gone, and there were a lot of hypotheses for that, but, basically, in this data set, there

was no evidence of spatial autocorrelation, and there was no need to use generalized additive models or spatial GAM models in the primary analysis. So, most of the paper doesn't involve GAMs, and there is really no use for it.

Now, Arden and I talked, and Arden said you know, Rick, I love these spatial models so much. So, why don't we, as part of our sensitivity analysis, why don't we put it in and see what happens?

So, there is a figure in the paper that includes a spatial GAM of the sensitivity analysis, and if we added this surface in, even though we don't need to, it turns out the PM<sub>10</sub> effect was robust to that. Now, of course, if there is no spatial autocorrelation, this type of thing should happen.

I went back and reanalyzed the data with the stricter convergence criteria. Now, we don't have an analogue of...like natural splines in two dimensions. In fact, like I said, the reason that we found this error in the GAM was that we had a two-dimensional smoothing spline model, and we are comparing it to our LOESS smooth for space, and we are doing simulation studies to figure out the properties of these models. In that case, the correlation between these variables is much greater, and the underestimation of variance was much more pronounced.

We first thought that we did a programming error and all this kind of stuff, and eventually, we convinced ourselves that we didn't, and then we had to sort of reverse

engineer S-plus to figure out what was going on.

But, basically, in these models, the model contained only one smooth variable of x,y coordinates and air pollution, and like I said, in those simple cases, all of these problems really don't seem to exist. So, really, the spatial...we don't really need to do it, but even if we put it in, the air pollution effect is robust to that.

Now, enough said.

**DR. HOPKE:** Let me stick in the first question, because one of the things you said in there was that now one could properly estimate the confidence intervals, but it was computationally intensive. Can you give us some idea as to how computationally intensive? What is the penalty for doing the whole ball of wax?

**DR. BURNETT:** Well, to do the exact estimator, you have to do...I think it is the number of the smooth variables squared times the number of data points squared, the number of GAMs. So, if I had 1000 data points like a short time series, it is 1000 times 1000, so I would have to do a million regressions.

Now, there is an approximation that has been developed for this, that I haven't developed, but Trevor Hastie had this developed, and it only requires about  $n$  GAMs. So, if you had 1000 or 10,000 time series study, you would do about 10,000 GAMs. And that can certainly be done. The question I don't know is how many people have looked at this,

how many people have used it. Like I said, Lucas has even, in a little data set that he is working on, found that sometimes the...when you run all these GAMs, that if you run 1000 of them, they don't always converge. The way that it is done is you are estimating sort of very sparse quantities in that.

So, it is tractable but...and it may be doable, the approximation is doable, but I think that I wouldn't run out and just do it. I think we need to study the properties.

**DR. MILLER:** You mentioned the number, but what is the computational time? I am used to seeing solutions for CFDs that run for three days for a single solution.

**DR. BURNETT:** Well, that is about...

**DR. MILLER:** Why wouldn't you invest and take...

**DR. BURNETT:** Sure enough. No, like I am saying, but like I say, I don't want to invest in something that I am not confident about. So, I think once...if it turns out as we study this and it turns out to be a reasonable approximation to the standard error, then the computational investment is not big, you know. I don't think it is a problem.

**DR. HOPKE:** Francesca?

**DR. DOMINICI:** I just wanted to make a technical comment. He correctly said, pointed out, that



whenever you have only one small function, using the most stringent convergence criteria doesn't matter, and also, when you use GLM, going to the natural cubic spline doesn't matter, and there is really an answer why it is not happening, and the answer is that this procedure of the generalized additive model, if you have only one small function, there is a closed form solution for the algebraic function. So, there is no convergence going on.

So, I just want to make clear that if you have a model with just one small function, the convergence criteria does not matter, because there is nothing that tends to converge. It is just one formula which will give you the answer. So, that is all.

**DR. MILLER:** I had one other question.

**DR. HOPKE:** Sure.

**DR. MILLER:** You mentioned, as you were presented, that such and such was the optimal for the span for this data. What is the criteria that is applied for that optimality decision?

**DR. BURNETT:** Well, obviously, everybody is going to have a different criteria. What...the philosophy that I take is a multi-stage philosophy. The first thing I do is find the span or the number of knots and the number of degrees of freedom and the smooths the line such that the evidence that the residuals are white noise is great. So, if I start to overfit the data, I start to get sort of a ringing

in the data. So, I try to make sure that that happens.

I also try to make sure that I have at least accounted for some seasonal cycles, because sometimes, you can get very low event rates, very small cities, and it is very hard to see the seasonality, but I still believe that more people die in the winter than in the summer even if they are in a small town.

So, you sort of take the criteria that...I want to at least account for seasonality, and then, how much more than that do I want to account for I based on a white noise criteria.

**DR. MILLER:** I think the point you are bringing up is it is, in part, up to the individual investigator as opposed to an abstract physical criteria.

**DR. BURNETT:** Yeah, there is no...and different people will have different criteria, and then I bring in the weather models and say how many...how much smoothing should I do to, you know, my best predictive power like an AIC or something like that. So, everybody is going to have...I mean, there is, I don't think, any...if I had a room of statisticians here, I'd have 1000 people with 1000 different opinions.

**DR. MILLER:** Just also, Bill, Dan, for a 2 gigahertz clock with a gigabyte of RAM in the same way of our CFB models, we have gone from days now down to hours, and you'll get there, too. I don't know where it is, but...

**DR. BURNETT:** Well, I guess one of the issues is that if we are sort of obligated to do this kind of sensitivity analysis, it will take a lot longer to do a study, and how many journals are going to want to publish 10,000 regressions, you know.

**DR. HOPKE:** Petros?

**DR. KOUTRAKIS:** I would like to say that factor analysis and the source apportionment, how you do it, in some cases, it is going to go...our update is pretty soon...when people realize, you know, that if you use these methods, it can be very subjective. So, I hope that people are not so...going to repeat the same mistakes.

**DR. BURNETT:** We didn't, but I think the message is, you know, always, you know, be cautious and keep doing good work. And the other things is, I mean, I don't view this is as an...people say this is an error or this is a problem. I view this as an opportunity, because now we have an opportunity to understand more about what is going on and an opportunity to maybe do something interesting. So, I don't view it as an error.

**DR. HOPKE:** George? And identify yourself, please.

**DR. POOLSON:** George Poolson, NYU School of Medicine. I just wanted to point out that this question of concurvity and each correlation of the weather terms with pollution, we have examined this as a published paper that

Cazio and I have in the Journal of Exposure Analysis and Environmental Epidemiology last year but with respect to ozone. We showed pretty much what you are pointing out, that as the interaction of the weather term with the pollution term rises as you get inordinate amounts of intercorrelations of the beta, then, you know, the pollution effect...I have a feeling that in the NMMAPS modeling, that may have been a factor in why ozone is not as significant as it would otherwise have been, but I think that is worthy of investigation.

**DR. HOPKE:** Rich?

**MR. POIROT:** Just maybe an observation to begin with.

**SPEAKER:** Use the mike, please.

**MR. POIROT:** There's a little bit of wood smoke and bricks folks sent down to us a couple of weeks ago. One of the...natural background, by the way. One of the influences was an estimated 10 to 12 degree suppression in the temperature over a huge spatial scale. So, I guess, you know, just there could actually be an error, you know, in your central diagram for the effect that air pollution has on weather.

And the other observation was that it occurs to me there is almost a whole separate body of literature in workers that are very involved in trying to understand simply the influences of weather on air pollution per se in order to

tease out other aspects of causality, and it would be a really productive future gathering to get that group together with your group.

**DR. BURNETT:** And I think one of the things that is going to come out of this is that, you know, there was a debate a few years ago about weather, and I think that some of the analysis done at that time said well, it doesn't really matter how you do weather modeling, and we sort of moved on from there, and I think that is something that we obviously have to revisit.

**DR. HOPKE:** Okay, I think we have really got to move on here. Thanks very much, Rick.

Our next speaker is going to be Suresh Moolgavkar. He is going to talk to us about his reanalysis of the Three-Cities study.

**DR. MOOLGAVKAR:** First, I would like to make a correction for the record. I am a consultant to...and instructor...an association of trade organizations...I said that I am a consultant to a consortium of trade organizations, including the ASI, but my affiliation should read the Fred Atkinson Cancer Research Center and the University of Washington.

I knew that Rick would do a terrific job, and he did, and I think he illustrated really well the fact that statistics is a good servant but can be a terrible master. I really feel, now, that some of the literature in looking at

the epidemiology of air pollution and human health, what we are seeing is the tyranny of statistics, fancier and fancier statistical technology pursuing smaller and smaller risks and, sometimes, I think, at the expense of common sense.

What do the GAM problems have to say about the results of previous analyses? One of the things I would like to say is something that Rick also observed, is that if you are looking at actual numerical values of estimates, then it can make a substantial difference, but if you want to look at the pattern of results, if you want to look at results across a lag of zero find days, or if you want to look generally at the shapes of curves, then the new convergence methods, the more stringent convergence criteria, don' t seem to make all that much difference.

So, let me look at one example here of an analysis that I did in Cook County. This is a full air analysis with only one pollutant in the model but with time and weather controlled using GAM methods that have been used for the last half a dozen years, and this is with the less stringent convergence criteria. You can see the shapes of the exposure-response relationships between daily mortality for a lag of zero to five days, and you can see that these shapes seem to indicate that there is little influence of PM on mortality for a concentration of more than 50 Fg/m<sup>3</sup>. So, you can see that for each one of the days.

So, in order to look at this a little more closely, what I did was I looked at the analysis of total mortality in

Cook County restricted to days on which  $PM_{10}$  exceeded  $50 \text{ } \mu\text{g}/\text{m}^3$

3. The top line is GAM with the less stringent convergence criteria, and the second line is generalized linear models with natural splines. What I have here are the coefficients and the standard errors.

What you can see is that the background remains more or less the same. You have three negative coefficients, albeit there is a fairly sizeable change in the numerical value of the coefficient, but you have both negative coefficients at lag 1, lag 4, and lag 5, and you have positive although quite insignificant coefficients at lag 0 and lag 2.

I have done only limited exploration using natural splines and using the more stringent convergence criteria, but this is my impression, namely, that broadly speaking, the general pattern of results don't seem to be altered too much. What I mean is you don't pay too much attention to the numerical value of the coefficients.

But now, what I would like to do is to take an example, and I am going to be using NMMAPS of the bottom case in which, when you combine results from 90 cities across the nation, what you are doing is taking, quite literally, the estimates of the coefficients that you get and the standard errors. In other words, the first basic assumption is that the only variable you get is the statistical variable you get, although that we know that the variability from

non-statistical sources is like to be much larger.

So, let me illustrate here what happens when you look at the old NMMAPS analyses and the new NMMAPS analyses. As was pointed out by Dr. Greenbaum earlier, this line, the horizontal and vertical line at zero, indicates that, in this quadrant here, there are a number of cities that were at positive coefficients in the old analysis but have either negative or zero coefficients in the new analysis.

But what I am looking at here is really the key statistic from the new analysis plotted versus the key statistic from the old analysis, and the red line up there, the dashed line, shows you above the horizontal line are the cities that were statistically significant in the old analysis, and to the right of the dotted line, you have cities that remain statistically significant in the new analysis.

You can see that out of the 90 cities, only 2 cities remain statistically significant in the new analysis, and there is one city, Little Rock, that has a negative coefficient that is more or less statistically significant. It is statistically significant.

Now, this could be sampling variability, but is it? Why New York and Oakland? New York, of course, is a large city, has a large power, a lot of daily deaths, so, of course, what you see is statistically significant. You expect to see a statistically significant result there if something is going on.



But why Oakland? Why not Los Angeles? Why not Chicago which is right around here? And Chicago not only has a large population with a large number of daily deaths but also has daily monitoring of  $PM_{10}$ , whereas New York and Oakland have only every six-day monitors. So, why is Chicago not significant in this, and why is Little Rock negative and significant here?

Now, I think there is more to determining whether there is heterogeneity in the data than doing a simple statistical test. I think Dr. Schwartz indicated that the chi-squared test of heterogeneity is not significant in this case. Well, that test has very little power. In any case, I think that there should be other considerations that go into this than simply statistical considerations.

Now, one of the things I wanted to do, I was curious to do, even though, in the presence of such heterogeneity, I don't think I would attempt to arrive at a single estimate of coefficient for  $PM_{10}$  effect, I simply wanted to see what would happen if those two outliers and this third outlier were removed from the data set and if we did a hierarchical Bayes analysis on the remaining 87 cities.

In order to do that, let me explain how my analyses differ from the analyses carried out by Francesca and her colleagues. I have tried to understand what they have done, and Francesca has tried hard to explain it to me, and my understanding is that there were two distinct models

considered by the HEI group. They looked at the 88 cities. I think that Honolulu and Anchorage were not considered, because they couldn't be assigned to any one of the regions. That is my belief. So, they looked at 88 cities, they did a Bayes analysis, and arrived at an overall estimate of the mean distribution and also Bayes estimates for the 88 cities.

They did a separate what they called a T stage analysis in which they looked at 88 cities that were divided into 7 regions, and then they had an overall effect. Now, it is not clear to me where the overall effect of 0.21 comes from, whether it comes from this analysis or this analysis, but this is my understanding of what was done.

Now, I find, before I tell you what I did...oh, I can also tell you that when doing the Bayes analyses, the Johns Hopkins group used what are called conjugate priors, and that simplifies the computation power of the posterior distribution. That is simply a fact, and that is the reason that conjugate priors are often chosen.

I wanted to have completely flat priors, meaning I have absolutely no information on what is going on, so I chose to use flat priors.

Now, here is something that I thought indicated a problem with the algorithm, the MCMC algorithm, that Francesca was using. I originally thought it was a convergence problem, but Francesca assures me that it was not. But I find this result somewhat incongruous.

You see, the range of estimates for the regions is

wider than the range of estimates for the individual cities. So, let's take a concrete example. For example, the model that is used consists of 15 cities. If you look at the Bayes estimates of the PM effects in these cities, they range from about 0.223 to 0.271. But the regional mean, Bayes mean, in the Northeast is 0.4, and I just find it difficult to understand how you can take the same data set, using the 15 cities from the Northeast, get a range from 0.223 to 0.271 there, and have a regional mean that is 0.4.

I find that...I am sure that there are statistical models, especially if these estimates are derived from a different statistical model than this estimate, I am sure there are statistical models that you could find this using the statistical models, but to me, I find it difficult to interpret this result. To me, it flies in the face of common sense. I would have a difficult time accepting that.

Now, here is what happens when you remove the two positive outliers. So, here is my hierarchical Bayes in the blue. The blue color here is my hierarchical Bayes distribution of the means for the 90 cities. It is pretty close to what Francesca reports. She reports a mean of about 0.21. I report a mean of about 0.19, and the median is quite close to the mean, although I don't think I have quite reached convergence here, because I am using flat priors, and I am using the parabolas Hastie's algorithm to look at the marker for gene cargo runs, and with the flat priors, I think convergence might be a lot more difficult than with the

priors that Francesca used.

Now, here is the distribution of the means minus New York and Oakland, and you can see that it has shifted from 0.19 to 0.12. Dropping out the two positive outliers in the data shifts the mean quite a bit.

Well, what happens if you put in Little Rock? What if you put in Little Rock? I have not shown it here, but if you put in Little Rock...sorry. If you remove Little Rock also in addition to New York and Oakland, then, as can be anticipated, the curve moves somewhere to the right of this red color, and the mean is around 0.15. That is exactly what you would expect to happen.

Now, here are the original data from the NMMAPS 90 cities. This is taken from Francesca's web site. And you can see the estimates, together with their 95 percent confidence intervals, and you can see also quite clearly how nicely the confidence intervals increase as you go to smaller and smaller cities, because these cities are in order of their size here in terms of population. But you can also see a considerable amount of scatter in the data.

Now, here is what we get through the miracle of hierarchical Bayes. This is what happens. This is the new beta and the new confidence intervals from the NMMAPS data analysis, and I ask myself, is this too good to be true? I mean, can you actually take the kind of data that we have looked at, go through these Bayesian analyses, and come up with something like this? And even if...one can show

mathematically, and I can show mathematically, because I find the same thing. I can show mathematically or statistically that this is the kind of thing that can happen and does happen, but is this a result that we should put too much credence in without thinking about it?

So, let's look at the assumptions that go into these Bayes analyses. Oh, there is another way to look at this. If you look at, along the x axis...

**DR. HOPKE:** We are running over here some.

**DR. MOOLGAVKAR:** Well, I think everybody ran over.

So, if you look at the HEI analysis, that is, the new coefficients from the new NMMAAPS analysis, single-city analysis, they range from -2 to 2. I have the same range on the y axis, and you see the HEI Bayes p statistic, and you can see how clustered it is, and the fact there that you have two that are statistically significant in the single-city analysis, but as a consequence of hierarchical Bayes, you see how many cities are all clustered around the statistically significant data here.

So, I think what needs to be looked at are the assumptions that are made in these Bayes analyses. The first assumption that is made and was discussed in one of the original papers and in the JAVA paper in 2002, I think, was that the actual data are replaced by the estimated parameter and the standard error. And a hierarchical Bayes analysis

should be using the real data, but you can't use the data here, because it would be computationally extremely intensive. So, the data are replaced by the estimated error.

Now, it seems to me that in order to use this approximation that the asymptotic properties of the likelihood have to be reached, and you have to have a very good approximation, a quadratic approximation, of the likelihood. Is the normal approximation adequate? We don't know that in every city, particularly where the standard errors are large.

The prior distribution for two city specific slopes is normal.

Well, how sensitive are the results to these assumptions? Why not choose a bimodal prior which is a mixture of two distributions with one normal centered at zero and the other one positive. I think this is a perfectly reasonable prior distribution. It reflects perfectly that, in some cities,  $PM_{10}$  is a good marker of air pollution effects, and in other cities, it is a poor marker. That would be my belief.

Here again, the hyperprior is also normal, and the results appear to be somewhat sensitive to this assumption, because I have used a flat prior. So, even though with my flat prior, I see that the Bayes estimates for the cities are clustered together which is just a property of this procedure, still, they are somewhat less clustered together than if you do not use the flat prior. This is the HEI Bayes

D versus my Bayes D, and you can see that I have a much larger spread, and in my analysis, the original two cities that were significant remain significant, and two more are added to the statistically significant group.

But...I will stop here, but my conclusion from all this is that I think that with all these assumptions that go into Bayes, hierarchical Bayes analysis, need very, very careful examination, and now that we have the opportunity, because of this S-plus problem, to take another look at these analysis, I would hope that this is a very careful and very well considered look.

Thank you.

**DR. HOPKE:** Any quick questions? Roger?

**DR. MCCLELLAN:** It is really to Francesca and for Suresh. Are these...when we do the combination, is there a population weighting in terms of that...

**DR. MOOLGAVKAR:** There is a population weighting just from the standard errors. You see, the original data have been replaced by the maximum likely estimate, and it is standardized. The implicit assumption there is that the approximation to the likelihood is adequate, adequate enough, whatever that means.

**DR. HOPKE:** Yes, Jon?

**DR. SAMET:** Suresh, I would hope that you would be posing these comments provided you don't think there will be a response from our team. You know, I think this is

not the place to go into...nor the time to go into details, but I think, you know, having voiced these concerns here, I think we need to see them in writing so that we can respond to them for the record.

**DR. MOOLGAVKAR:** Well, I think you have most of them in writing.

**DR. SAMET:** Well, we are talking here about the presentations today to the CASAC panel. I think, to further the comments that you made on the transparencies that we should just make sure that they have the opportunity to respond to them.

**DR. MOOLGAVKAR:** Sure. I think most of them have already been...

**DR. HOPKE:** Okay, our last speaker is, then, Mitch Klein from Emory University. Are you transparencies or a computer man?

**DR. KLEIN:** I am not either, but I'll use the computer.

This is just what you are in the mood for, I know, more GAMs, but I'll try to be very quick, because when I started to do assimilations this message wasn't out there. I didn't even know what the results would be, but everybody has talked about it. I think now there is a general agreement that GAM variance estimators do seem, to the available software, specifically, S-plus and SAS, do underestimate the true variance.



What may be of interest...we have talked about multi-city studies, but what is the effect in one city, namely, Atlanta? As a component to the study of particles and health in Atlanta, we have been attempting to link daily measures of air pollution with daily counts of emergency room visits. So, generally, as you have seen these models several times today, with these daily count outcome models, they are typically modeled as Poisson with a log link, and the model contains a quadratic term for air pollution, and some function of time trend, some function of meteorology, and then other covariants.

What separates a GAM from a GLM are the way those functions are modeled. This time-trend meteorology in the GLM are functions...models functions of regression parameters. With GAMs, any GLM is a GAM, but what makes GAM different is they allow these nonparametric smooth functions such as smoothing splines or LOESS.

Now, when we ran GLMs and the corresponding GAMs, we did notice the parameter estimates were pretty close, but systematically, there was a difference that the GAM differences were larger or smaller. So, the question of interest to us was, are they smaller because they are really that much more efficient, or is it the case that the variance estimates, the air pollution parameter estimates, are underestimating the true variance?

So, to help answer this question, we have done a series of simulations, and I will quickly outline the methods

used. It shows...this is just one scenario. I want to make clear that it is several scenarios, and it is all combinations, a lot of combinations...should have absolutely an emergency department outcome. Then, we used the data from our Atlanta study to fit a specified model, and this acts as the true model, the underlying model that generates the data.

From this model we generate quasi- distributed outcome times series using the fitted values as the Poisson means. In other words, we look at our fit...generally, we never know what the Poisson mean for a given day is. We think it...we are pretty sure it is a function of season and a function of time, it is a function of air pollution and what not, but this time, we are saying that is the truth, that we observed the underlying mean.

We repeat this process 1000 times, and then we analyze the generated data using GLMs and GAMs. So, we do this 1000 times. We have 1000 parameter estimates.

We can then look at the variance of these 1000 parameter estimates, and that can act as a proxy for the truth. That will be our gold standard.

We also have 1000 or along with 1000 parameter estimates, 1000 estimated variance estimates for the parameter, and we can look at those and take the average of them and compare that to our proxy for the truth.

Now, we hope the ratios between those two estimates are approximately 1.

So, scenario 1: Scenario 1, the data was generated

using B-splines. We are modeling the CBD and PM<sub>2.5</sub> association as they were in Atlanta. We are using monthly knots or equivalent degrees of freedom for the GAMs. B-splines and N-splines which are the first two, their GLMs, the smoothing splines and LOESS to GAMs.

So, using B-splines as the truth, we had a parameter value of 0.0277. Exponentiate that, and you get the risk ratio which is about 1.28. If we had 1000 estimates for each of those four scenarios, so there are 4000 estimates in all, and we take the mean of the estimates, we can compare that to the truth. Next is the difference. Then, we can take the standard error of the 1000 estimates which is our proxy for the true standard error...sorry, that's the following one, standard deviation of the estimates. First, if we have 1000 standard errors, we can take the mean of that for which the first one would be 1.0122. Then, we can take the standard deviation of the estimates. We can take the ratio of those previous two columns, and that ratio should be close to 1. And, finally, we can get 1000 estimated 95 percent confidence intervals, and we can see the proportion of time that those confidence intervals cover the true value which we know.

So, the bottom line is if we look at the ratio of this estimated standard error to our proxy for the truth, for the B-splines and N-splines, they are about 98..between .98 and .99, and for the GAMs, the S-splines and LOESS, it is about 83 or .83. And the coverage in the confidence

intervals is very close to .95 which we would like, but for the GAMs, the 95 percent confidence intervals are actually 90 percent confidence intervals, and this is actually one of the highest scenarios, the most favorable scenarios for the GAMs.

Let's look at another scenario, very similar. Same pollutant outcome. The only difference is this time, the underlying model was generated using LOESS. So, now, we are saying LOESS is the truth. So, what happens if we do the same four analytic models?

Well, again, focusing on the ratio which is the second to last column, the standard errors, the ratio is about 1 for the B-splines and N-splines and 84 percent, .84, for the GAMs. This is just one example, so I don't want to make too much of this sort of bias, but it is interesting that the GLMs in this example fit the underlying model being LOESS actually better than the GAMs do in this example.

So, here is actually 15 more scenarios, just what you want to see. Three column outcomes. The underlying model for all of these is LOESS. And five different types of analytic models. So, we are going to have 15,000 models just for this page.

The first three columns of numbers are GAMs; the last two are GLMs. Again, the ideal ratio would be 1.0. That means the standard error estimates should be working right.

You can see for the GAMs that it goes from a range of 0.769 to, I guess, 0.88, and for the GLMs, it seems to

work pretty well, close to 1. So, that is the bottom line, actually.

Just to get the...this is just one example to give a feel for the convergence criteria for  $PM_{2.5}$  and CBDs in this data. So, default to  $10^{-3}$  up to  $10^{-15}$ , a seven decimal accuracy which is really more than you need for parameter estimate. Looks like it is okay by  $10^{-9}$ .

Now, in this case, it does go down. In our data, for particles, most of the times, it went down, but in our data...and I would be curious if other people find this for ozone, using the more stringent convergence criteria, our average parameter estimate went up.

This was rather nice change. Sometimes, it was stronger. Generally, the convergence criteria, using the stringent ones made the answer closer to the GLM, whether it went down or up.

Model misspecification, a big topic. Now, simulations are kind of artificial, because you generate them knowing what the truth is. In real life, most models do not know the true model. It is actually quite fortuitous if we happen to come upon the true model.

So, we assume an underlying structure. Assume sounds a lot better than pretend. So, I am actually going to show you just three examples. There are a million misspecifications you could do, but one is, what happens if we speci...if in the underlying model you have more degrees

of freedom than the analytic model? Then, flip it around. What happens if we overparametrize the model and have more degrees of freedom than the underlying model? And the third scenario is, what happens by shifting knots in a GLM?

So, in this case, the underlying model was modeling with seasonal splines, seasonal knots. Is that right? No, I am saying that wrong. The data generated was LOESS. The data was analyzed using seasonal appearance, but the underlying model was monthly knots. So, we have too few degrees of freedom.

And in these four scenarios, the true value of 0.028, it was underestimated using all these methods. So, but the ratio of standard errors, it didn't affect that. It was still okay for the B-splines and natural splines. The confidence interval suffers, of course, when you get a bias truth, because the confidence intervals won't come to the truth as many times even if you have the right variance.

And now, it is misspecified the other way. This time, the data is generated using a knot every season but analyzed with monthly knots or equivalent degrees of freedom in the GAM case. In this case, there is much less bias, if any. And I don't want to generalize this or something, but in this situation, it is better to overparametrize than underparametrize, but, of course, when you overparametrize, you run into the risk of controlling for the air pollution effect, and you don't want to control for that.

So, there isn't much bias. Again, the standard

errors show the same pattern. Standard errors consistently show the same pattern.

One final misspecification, now, the data was generated here by placing a knot using B-splines on the 21st of each month. So, the question is, what happens if you shift it to the 7th, shift it up to the 14th, the 21st, or 28th? Now, there are a lot of different ways you can change knots, but just in this scenario, it was very robust to a shifting of knots.

I'd like to say one word about the default for placing knots. The default is quantile, but I would recommend actually specifying the knots so you are aware of exactly which day the knots occur.

I think there are several good reasons. One is just it is nice to know exactly what your model is, where the knots are. The other is with missing data, quantiles in the data does not necessarily correspond to quantiles in time, if that is what you are interested in. So, if you actually decide...someone says oh, the PM<sub>2.5</sub> measures really aren't bad in the week of February 10th, set them to missing, well, that won't affect the parameter estimates, but if you use the default criteria just with quantiles and then rerun it, you are actually changing the model. You are placing the knots at different spots.

Also, if you have, you know, you have, let's say, different beginning points and end points, a 29-month time period, if you just use quantiles, in different calendar

years, you might have the knots at totally different places.

That is my recommendation. And my final summary slide is this. Big surprise. The variance in GAMs are underestimated using the standard software, and in our simulations, in Atlanta, it was a factor of 0.75 to 0.90 as a general range of the standard errors.

I should point out, in Atlanta, the seasonality is not nearly as strong as in, say, Canada, so I could see in other cities, it could be stronger than that, and that impacts the validity of the confidence intervals and statistical test.

Thank you.

**DR. MILLER:** Is the take-home message in general from the GAMs that what you are thinking are 95 percent confidence intervals are more likely 90 percent confidence intervals?

**DR. KLEIN:** I can't say it that strong. I'd say what a 95 percent confidence interval is what you think are less than 95, and in that example, it was 90 percent. There are other examples where I have it in the low 80s, and this is just Atlanta. Because the standard error is underestimated, it will definitely be less than 95 percent.

The other thing about confidence intervals, it really depends on what the true value is. So, that is another factor.

Yes, sir?



**DR. BURNETT:** Did you have weather in those models?

**DR. KLEIN:** Yes, I have weather is 5 degrees of freedom for both dew point and mean temperature.

**DR. BURNETT:** So, a number of those simulation results look like to me that the smoothers gave you unbiased...mostly unbiased estimates of the air pollution effects.

**DR. KLEIN:** You mean the nonparametric scheme?

**DR. BURNETT:** Yeah, even with weather models, you know, weather variables in there. So, there is not a lot of concurvity in weather?

**DR. KLEIN:** Yeah, I think in Atlanta, there is much...I know there is much less. So, I think that probably is a reason compared to the Canadian models.

**DR. GREENBAUM:** One thing I was confused about. It looked like in your...on your parameterized model, that is, too few knots in your analysis here, that the bias in the data was down, and this seemed to be different from what we were seeing in NMMAPS and other analyses where, actually, we were increasing the number of parameters but the estimate was decreased. Is there a reason for that?

**DR. KLEIN:** This is the way that...this is one example, so I don't want to generalize at fast pace. So, I don't know.

**DR. HOPKE:** All right, thank you very much.

What we want to do at this point is move to the public comment period. We will do this without a break, so any of you who need to duck out, duck out as needed. Don' t everybody go at once. So, we have allocated ten minutes each, and we will be quite rigorous with that so that we can get done here at a reasonable time today.

So, we will just go down the list and get started, and as I say, ten minutes each. We want the speaker, the next speaker, to come up into the on-deck circle so that we can move rapidly from one speaker to the next and move...keep things moving right along.

So, our first presentation is from the American Lung Association, and that will be given by Deborah Shprentz.

**MS. SHPRENTZ:** Good afternoon. I am a consultant to the American Lung Association.

Today is the fifth anniversary of the establishment of the NAAQS for fine particles which marks the five-year deadline for the completion of the review under the Clean Air Act. This milestone is especially critical because of the commitment made by EPA Administrator not to enforce the fine particle NAAQS until the standards had been reviewed. Now, with the deadline for the completion of the review upon us, we find ourselves still more than a year off from a final rule.

Meanwhile, EPA and the States have not begun the

process of implementing control strategies that could prevent the 15,000 premature deaths each year that EPA estimates are associated with PM concentrations above the level of the standards. Each delay in the completion of the standards review process imposes a high cost on those that suffer health effects from breathing particulate air pollution.

In order to maintain momentum, we urge CASAC to reach closure on Chapter 1 through 7 of the Criteria Document which are completely unaffected by the recent software problem.

The message we heard this morning is that the software error does not affect the major conclusions of the Criteria Document. Reanalysis of the NMMAPS study changes the quantitative estimates, but the major conclusions remain the same, that is, there was strong evidence of association between acute exposure to PM<sub>10</sub> and daily mortality, particularly from respiratory and cardiovascular causes, but this association cannot be attributed to other pollutants or to weather.

The software error pertains only to certain recent time-series studies, and reported error has no effect on the results of the landmark long-term studies of particles, the Harvard Six-Cities study and the American Cancer Society study, both of which were reanalyzed in depth and upheld in 2000. These studies found that prolonged exposure to particulate air pollution significantly increases the risk of dying from cardio- pulmonary causes.

Importantly, since the long-term studies provide the basis for the large risk estimates associated with PM air pollution, these estimates remain unchanged by recent developments.

We would like to see EPA establish a process and an accelerated timetable for the completion of the reanalyses and the review of the results and to establish firm deadlines for the completion of the NAAQS review process.

On June 20th, the California Air Resources Board unanimously approved lowering the annual average standards for PM<sub>10</sub> from 30 to 20 Fg/m<sup>3</sup> and establishment of a stringent new annual average standard for PM<sub>2.5</sub> of 12 Fg/m<sup>3</sup>. CARB wisely decided to move forward in adopting new annual average standards, because the studies upon which they were based were not affected by the NMMAPS software error.

California is a leader in air quality protection, and the American Lung Association strongly supported the California standards for PM. EPA's staff paper much include options for more stringent annual average standards such as those recently adopted in California.

CARB was also poised to approve a stringent new 24-hour standard for PM<sub>2.5</sub> of 25 Fg/m<sup>3</sup>, not to be exceeded. California will take up consideration of the 24-hour standard when the review of the effect of the software error in the time-series studies has been completed.

Importantly, California employed the

not-to-be-exceeded form for all of its PM standards, both the 24-hour and annual average. The form of the standard is as critical as the level in dictating health protectiveness.

The 98th percentile form of the current 24-hour PM<sub>2.5</sub> standard allows seven exceedence days each year. This negates the purpose of the 24-hour standard, that is, to prevent the health consequences of high daily concentrations that are not controlled by the annual average standard. We note that EPA's exceptional events policy ensures that wildfires or other natural events are excluded from calculations of nonattainment.

The American Lung Association believes that the current 24-hour PM<sub>2.5</sub> standard is not protective of public health. EPA's last staff paper cited 65 specific effects estimates from U.S. and Canadian studies associating daily increases in PM<sub>2.5</sub> with total mortality, cardiovascular mortality, respiratory mortality, hospital admissions for cardiovascular causes, respiratory causes, COPD, and asthma, and with respiratory symptoms.

The vast majority of these associations demonstrating a distinct short-term effect are statistically significant. Most of the studies reported effects at levels below the current standards and in the range of contemporary concentrations in many U.S. cities.

According to preliminary analyses of PM<sub>2.5</sub> monitoring data for the last several years, only a handful of

areas will exceed the 24-hour standard, and these areas will all also exceed the annual average standard. The monitoring data indicate that major metropolitan areas that attain the annual average standard will continue to have high 24-hour concentrations at levels clearly indicated to be unhealthful in many studies. Thus, the annual average standard, while lowering distributions, is not sufficient to protect against short-term effects.

The evidence is clear that the current 24-hour standard is ineffective, and the staff paper needs to include options for strengthening the form and level of the 24-hour standards.

The case for strengthened air quality standards for PM grows stronger each week with the publication of new studies on the health effects of particulate air pollution. In recent months, there has been an explosion of compelling new studies linking particulate air pollution with lung cancer, deaths from cardiovascular causes, vasoconstriction, atherosclerosis, stroke and heart attacks, lung inflammation, asthma, and reduced growth in children's lung function, and retention of particles in the lungs and translocation of particles to the brain tissue. The need for action is strong and urgent, now more than ever.

With a \$200 million investment in PM research, great progress has been made in addressing the questions posed by the National Research Council. Recent research has addressed each of the major industry criticisms of the

science of PM. Given the strength of the new science, we find the tone of the Criteria Document to be overly cautious and equivocal.

A number of concluding sentences in the integrative synthesis strain to point out continued uncertainty. We believe that the role of the Criteria Document is to say what is known. This is not a research needs document.

In summary, we urge CASAC to reach closure on chapters 1 through 7 of the Criteria Document. We urge EPA to establish a process and schedule for completion of the reevaluation of the time-series studies and an accelerated time table for completion of the NAAQS review process. And we look forward to the development of a staff paper that includes options for more stringent annual average and 24-hour standards such as those recently adopted or under consideration in California.

Thank you.

**MR. FLAAK:** Thanks, Deborah.

We have an adjustment in the schedule for the speakers, but before we get to the next speaker who will be Ron Wyzga on behalf of EPRI...he is switching with Rebecca Klemm...I just wanted to ask how many other speakers might be willing to, if we have to run into tomorrow, I have one volunteer for tomorrow morning. Anybody else? Fred? Okay, thank you. Just in case. We may not need to do that.

Again, handout materials, I would appreciate getting those, and if you use overhead slides, if you don' t

have copies of those today, we can get those made while you are here.

And one last thing. For folks taking breaks or stepping out of the room, it would probably be better if you did not use the back door, because every time someone comes in and out, that rattle sounds through the whole room. Use this doorway here. This one seems to close quietly. I appreciate that.

Thank you.

**MR. WYZGA:** I want to talk about the other cohort study that never seems to be mentioned and I've talked about it before, which is a study that we have been involved in with a group at Washington University, the methods study. You have seen most of these before. I will just sort of highlight. This is a study of 70,000 U.S. veterans who were treated at Veterans' Administrative hospitals. They have been followed since the mid 1970s, and there are some characteristics of the cohort listed here, and you can look at that at your leisure.

There are a couple of differences...

**SPEAKER:** Could you try to focus that more, please?

**MR. WYZGA:** Sure. There are a couple of differences from some of the other cohort studies. One is that we looked at peak ozone and peak carbon monoxide data rather than the annual average by taking the 95th percentile



of hourly the ozone data and hourly the carbon monoxide data.

Because we were following the group for a period of over 20 years, we thought that it didn't make any sense to choose one value of air pollution for that 20-year plus period and look at total mortality for that period but to break it up into different groups so that we would not be in the position of looking at deaths that occurred before we measured the pollution or looking at deaths that may have occurred 20-plus years after the pollution, so we broke it up into different periods.

We looked at county-level air quality data rather than data for an SMSA, and, also, we looked at ecological variables that were included at the zip code level.

The point I want to make, and I have shown this before...I only have data for Cincinnati that go back to the...up to the 1940s and 1950s, but if air pollution is impacting the health of the cohort, we don't know to what extent it might be some of the earlier levels which were really much, much higher than some of the later levels. We are talking about people who were dying post-1970. They were younger in the period pre-1950. The question is, did those values impact their health or lung development? And the answer is we don't know, and that is one of the things we wanted to look at in some of this study.

So, the basic design of the study was one where we had four different air quality exposure periods, and we looked at mortality in three different time periods, and we

looked at the association between the different elements in this matrix so that we could find out whether or not the early pollution seemed to affect deaths more or less simultaneously or deaths at a later period in time.

When we did this, the results for particulate matter were strikingly negative. We don't know why, but they were negative. Whether we looked at PM<sub>10</sub>, PM<sub>2.5</sub>, PM<sub>15</sub>, the results were negative.

What was significant, however, were some of the gases, in particular, NO<sub>2</sub> and ozone, and when we basically tried to look at ozone and NO<sub>2</sub> together to see before we could tease out of the model, the stronger effects came out to be ozone rather than NO<sub>2</sub>.

Now, the study is described in the Criteria Document, but it is criticized and dismissed for several reasons, and I want to sort of address these one by one. One is it said it should be given less attention, because the cohort is an all-male cohort; secondly, that it included a large number of former smokers; that it has a wide range of exposure mortality periods...I tried to explain why we did that...and it also said because it is a smaller study population than some of the other studies, and this is certainly not true for the Harvard Six-Cities study which has a much smaller study population than this study, although this study is smaller than the American Cancer Society study.

If we look at the gender issue, we looked only at

males. If you look at the most recent study of Pope et al, the lung cancer effects are significant for males, not for females. If you look at the Harvard Six- Cities study results, the relative risks are higher for males and not females. In the American Cancer Society cohorts, there was not much of a difference between males or females.

So, I submit that the fact that we looked only at males does not bias the study in terms of looking at the healthier sex.

The second issue was one of looking at smoking status. I guess two things here is that we had about 80 percent of this cohort were former smokers. This is a quote from the Criteria Document, basically saying that they saw the strongest evidence of PM effects in current smokers and in non-smokers. In the basic ACS results, the most recent lung cancer results, the results were similar for smokers, ex-smokers and current smokers. Also, we looked at the data, quote, basically looking at both smokers and non-smokers, and we didn' t find any difference in the air pollution coefficients whether we looked at the smoking population or the non-smoking population.

The principal purpose of the cohort was not one trying to look at air pollution but one trying to look at the impacts of blood pressure and medical intervention on health outcome in this group of veterans. So, we had detailed blood pressure variables. The investigators at Washington University felt it was very important to include these

variables in the model, because they do very much sort of explain health outcome and mortality outcome in particular. One finds a very complex relationship between blood pressure and age.

So, there was some criticism...there was criticism that by including blood pressure as an independent model that perhaps we were explaining away air pollution, because air pollution was impacting the blood pressure which then was impacting mortality. So, it was suggested that we...so, what we decided to do, then, was to look at...break up the cohort into two groups, and we basically looked at those that had diastolic blood pressure less than 95 mm mercury and greater than 95 mm mercury.

We found that, you know, depending upon the pollutant, there were differences. There was, if anything, a slightly greater effect for ozone when we looked at the people who had lower blood pressure. For particulates, it was very mixed. It depended upon the year combination you looked at, but, by and large, there wasn't much of difference whether we looked at people who had the higher or the lower blood pressure levels, and for PM<sub>10</sub> for the respective time periods, again, there was relatively little difference.

So, our conclusion in looking at this is that we didn't see that including blood pressure per se in the model had any impact on the results, and I will show you something else that was done in a moment.

The other comment that was made is that, you know, we had thrown in some, I guess, more explanatory variables, independent variables, in the model that some of the other studies, and it was criticized and said that perhaps we should consider more parsimonious models.

So, we did that. We basically looked at models that had fewer variables, and let me try and explain what we did here. Let me basically...so, we did it for the different time periods. Here is an example that is sort of very typical when we looked at the impacts of 1982-1988 ozone on, basically, the contemporary mortality period, and what I have in the bottom graph is the ozone coefficient, and on the top graph, I have the...I cut the information criteria, and, basically, you want to...basically, the optimum model is the one that gives you the lowest number here.

What you find in the bottom here is I have our baseline model which shows you what the coefficient was, and then I serially deleted some of the explanatory variables. For example, height was one that the literature suggests that taller people live longer than shorter people. When we took that out of the model, we had a slight difference, but it really didn't affect things very much.

Next, you have got body mass and age interactions. That didn't affect the results very much. We next took out the age and blood pressure interactions. Very little effect. Took out the blood pressure itself, took out the blood pressure diastolic and then diastolic variable itself, and

you see that there are differences, but they are not very great in terms of the coefficient of the model, and, in fact, the Ikaki information criteria suggests that this may be the optimum model, and it doesn't vary very much at all in the results that we find when we...the results of the full model specification.

If I look at the same data for  $PM_{10}$  ...this is a model looking at contemporary  $PM_{10}$  ...we find that as we delete more variables, we get some increase here, but, again, it is not that large an increase in the coefficient of the  $PM_{10}$  variable, and, in fact, the optimum model, again, seems to be the one that gives us the lowest estimate of  $PM_{10}$  coefficient.

So, my concern is...and I think it is...I don't know why we get these results, why we found ozone, why we found  $NO_2$ . We didn't find  $PM_{10}$  or  $PM_{2.5}$ . My concern in the document...and I'll give you another example...is that results that are unpopular that somehow don't seem to fit the mold may be downplayed or ignored, and I am really concerned about that, and I will give you another example.

This is a rather esoteric one, esoteric area, but it is illustrative of what happens. There is a paper...in fact, we had sponsored some work by Chris Murray and Charles Nelson addressing the harvesting issue. It had an interesting approach, a different approach, and it reached

some different conclusions from some of the other papers that have addressed this issue, and, in fact, it concluded that harvesting was really quite important.

I actually hand-delivered this paper to a representative of EPA in January of 2000. We sent it to EPA physically later in the spring of 2000 and 2001. It wasn't cited in the last draft of the Criteria Document. The principal author of the paper sent a very angry letter to EPA saying why isn't this even considered and here are my arguments. This time...and, in fact, we sent a letter. This time, it is listed in the references, and it is listed in a table that is totally...has nothing to do with the topic whatsoever, but it is not really discussed in the current Criteria Document.

Now, I don't really know whether this...I don't want to make any judgment whether this is the definitive paper, but my concern is it raises an issue that there is another way of looking at something, and I think the Criteria Document deserves to treat...to take all this information and to examine it and to try and make some judgments as to where we are.

There is a second paper in the literature that is published in a more obscure journal by Richard Smith who is a professor at the University of North Carolina, and he uses a model very similar to the one used by Murray and Nelson, and that is not even cited in the Criteria Document, although I think that paper has been called to the attention of EPA.

In addition, I have a new staff toxicologist who quickly read the toxicology chapter, and she immediately identified seven key studies that weren't listed in that chapter, all of which were...or most of which tend to have negative results or show effects only at very, very high levels. None of them is cited there, and I am concerned about that.

So, I guess if I have a bottom line, it is one that I urge you, CASAC, and I urge the Agency to really include and address and discuss all of the relevant studies, not simply to select the ones that may fit some kind of a mold, and, secondly, that we don't take...throw out the negative by different results, that we basically try and examine the results, see how they fit in, and try and give us a better understanding of what is going on. I suspect...I think we all agree that the more we dig into this, whatever is going on, it is something that is very complex, and very simplistic attempts to address these issues are not going to resolve this problem.

Thank you.

**MR. FLAAK:** Our next speaker is Dr. Michael Goodman who is incorrectly listed as with Hunton & Williams. He is with Hecksbar, Incorporated, and he is representing a utility air regulatory group. Need the overheads? Thanks.

**DR. GOODMAN:** Good afternoon. These comments were put together by my colleague, Michael Halperin,



and by me, but Dr. Halperin chose to go on vacation to Greece, so I will take this opportunity and speak for the both of us.

Fundamentally, our review of the most recent version of the Criteria Document revolves around five issues. These include errors in statistical analysis of the kind that have been discussed today, and there is no reason to go into it any further; lack of study of spatial pattern that we identified in the beginning of the NMMAPS study; confounding by copollutants; the health impact of unspecified variables; and, finally, the new data indicating generally more of a biogenic particulate matter.

With respect to the errors in the NMMAPS analysis, I really don't want to dwell on it today. For me, as somewhat of a newcomer to this area, looking at the results on left-hand side, one would say yeah, but there may be a story that is emerging, but if I did not know the preexisting history and just looked at the right-hand side, I would say I don't know what to make out of the results. Two cities out of 88 or 90 show statistically significant results, 3 out of 7 for only one set of analyses showing statistically significant results by region just does not seem particularly convincing.

If one were to take all cities and rank them, an analysis, we are thinking, somewhat similar to one of the previous speakers, one would take the data estimates, divide it by standard error, and then rank the cities according to

that T statistic. As expected, New York and Oakland would end up on the right hand to the most extreme locations, because these are cities that showed the strongest results.

One would expect, then, that neighboring locations, cities in the neighborhood of these two, would show...would rank somewhere close. This is not the case, however. We have New York and then Jersey City and Newark somewhere in the middle. With respect to Oakland which ranked second after New York, you have the neighboring areas in California, Sacramento, Stockton, and Modesto that are actually on the left extreme of the spectrum.

Another comparison would be, say, Toledo, Ohio and Cleveland. I am trying to find it. Right here. Quite opposite. Although they appear to be grouped together in terms of region, they seem to indicate very different findings.

With respect to confounding, it is clear that the authors of the NMMAPS, after reanalyzing the results, concluded that copollutants do not matter in affecting the results. However, there are a number of studies that seem to find very different results.

For instance, NO<sub>2</sub> seems to have a very strong effect on PM mortality associations in the APHENA study. Moreover, PM by NO<sub>2</sub> ratio seems to have a statistically significant negative impact on mortality.

Similar observations, although with different

copollutants, are true of the recent Brazilian by Kuntz and Flowers, et al where the effects of particulate matter and childhood mortality were evident in single pollutant models but in single pollutant models only.

Other studies would include a Canadian study by Stieb, et al, and this fact was actually acknowledged by the PM CD, indicating that this study shows no independent effect of particulate matter.

Similar results were tentatively reported by Burnett, et al in 1999, but the discussion of the multi-pollutant analysis result is no longer present in the document.

There was some important discussion of this issue that was present in the previous version of the document that appears to no longer be included in the third draft.

Another important issue is the impact of unspecified variables. If one were to look at some of the findings of the Six-Cities study and ACS study, what is striking is that the relative risk estimates for the association between particulate matter and mortality are very different. They differ by different demographic characteristics.

For instance, as noted earlier, there is also only positive for male, statistically significant positive for male, but not for females. For females, there is a difference by marital status. A particularly striking difference is by level of education where less than high

school education is associated with relative risk of 1.45, while more than high school education showed no additional risk at all.

Possible explanations would be there is a correlation with other unaccounted factors such as, for instance, health-related behavior, or possibly some analytical errors that could produce such unexplained findings.

It is also important to note that seasonal variation may have...may be explained by biogenic particulate matter such as pollen, microorganisms, mold. These certainly may have an impact on a number of outcomes under study and may have an impact on morbidity and mortality. These are just five studies that reported findings consistent with that explanation.

Again, it is somewhat disappointing that, compared to the previous version, some of the important text discussing this issue is missing from the third draft.

With respect to our overall recommendations for the next Criteria Document is that, for obvious reasons, the results from NMMAPS and other studies should be corrected and the Criteria Document revised accordingly. It is our feeling that multi-pollutant models should take priority over single-pollutant models and use the statistical approach. Finally, one needs to...not finally, but thirdly, one needs to explore the differential relative risk estimates by various demographic and educational characteristics, and,

finally, the role of biogenic material needs to be explored as another potential confounder.

I have to add from...say from my previous life in the emergency room, I knew exactly when to expect a rush of admissions and, potentially, death, and these are usually explained by sort of infectious disease epidemiology that is out in the community, whether it is an RC outbreak in children or a flu epidemic. These things may be important explanatory factors that could confound the results.

Thank you.

**MR. FLAAK:** Thank you very much. Our next speaker is Allen Lefohn. Dr. Lefohn?

**DR. LEFOHN:** I am Allen Lefohn. I am from Montana, and most of these comments will be on my web page starting next week. They have been submitted in hard copy.

Comments I am going to make today reflect the opinions of myself, Professor Paul Switzer from Stanford University, and Dr. Wayne Ott, also from Stanford University.

The bottom line of Paul Switzer, after reading Chapter 8, was that a multiplicity of cities does not guarantee that there are not important model deficiencies in the common model and the statistical methods relied upon in the Criteria Document. Because of the deficiencies in the Criteria Document, we cannot draw comfortable conclusions regarding the circumstances and magnitude of ambient PM health effects or whether reported PM health effects are

positive.

In his comments, he provides recommendations for improvements to the modeling effort.

Bottom line problems, unexplained heterogeneity of PM health effects estimates, enforced linearity of exposure-response, meaning the lack of a threshold issue by E. There is evidence that Dr. Switzer believes, based on the published literature, that such a threshold may exist. If so, that presents serious problems to the model.

The issue of copollutant confounding and enforced additivity in the analysis model itself. Once again, this information is provided in 15 pages of comment that Dr. Switzer had.

The importance of a threshold, the existence of a biological threshold has potential impact on the level of the PM standard that has been selected and evidence of reduced PM mortality resulting from reduced PM air pollution. The epidemiological models assume linearity and no threshold, and the assumption of no threshold plays an important role in much of the supporting material that the CD relies upon.

The existence of a threshold or nonlinear effect would call into question many of the published results cited in Chapter 8. In other words, we believe that the issues go well beyond just the existing model problems that were talked about today. There are also very, very important issues that we feel need to be addressed in addition to the important issues that have been discussed.

Spatial gradients, Chapter 3 concludes that fine particle concentrations are less spatially homogeneous in many areas than have previously been assumed. Chapter 3 states that although PM<sub>2.5</sub> concentrations may be highly correlated between sites, the concentrations themselves may not be spatially uniform, i.e., correlation is not a measurement of spatial variability. Work by Ito, et al and current work by Dr. Paul Switzer's research group show that the presence of spatial variability within a study area results in varying mortality estimates.

In the CD itself, there is a lack of consistency. In the executive summary, the CD states analysis of recent data from the PM<sub>2.5</sub> monitoring network show reasonable site-to-site correlation among cities. This indicates that, in such cities, the concentration at the air monitoring site or the average of several such sites will provide an adequate rep of the concentration at a sited home, i.e., focusing on the correlation instead of the absolute differences within the monitors.

This line of reasoning occurs in Chapters 5, 8, and 9. There were no scientific data presented to show the correlations instead of absolute differences among monitors is what is important.

It is important to note once again to emphasize that a high correlation coefficient between monitoring sites does not necessarily mean that the monitors' absolute values

are the same. High correlation only indicates that the sites are increasing and decreasing at the same time.

The ramifications of the spatial variation that has been pointed out in Chapter 3 now, there is exposure misclassification. Average concentrations cannot be used as a surrogate for community personal exposure where you have this type of variability occurring, and the existence of a threshold effect that is coupled with the variation in concentrations within a study will affect the predictions.

The variability of 24-hour average background PM, estimating background PM concentration is important for the EPA's health risk analysis that will be going on in the staff paper. It is important for the risk assessment to use the quantified 24-hour background PM levels at clean sites. Without an adequate characterization in Chapter 3 which does not exist at this point regarding the clean western sites, there will be insufficient information in the CD from which the staff paper can draw for the Agency's risk assessment.

In other words, please leave Chapter 3 open so that additional information can be put in concerning background.

Chapter 3 states that peak 24-hour average natural background concentrations may be substantially higher than the annual or seasonal average natural background concentrations. As an example of the available data we have characterized, I have characterized background particulate matter in using data from the 14 approved network sites or from 14 sites. There are many more sites than that.



For the approved sites, we characterized the annual mean for  $PM_{10}$ ;  $PM_{2.5}$ ; coarse, the annual mean of  $PM_{2.5}$  sulfate and nitrate; and the annual mean of  $PM_{2.5}$  elemental carbon and organic carbon; and the percentile distribution of the 24-hour concentrations by year for  $PM_{2.5}$ ,  $PM_{10}$ , and  $PM_{10-2.5}$ , in other words, the distributions of the 24 concentrations and not just the annual averages, not the smoothing; percentile distribution also of the 24-hour concentrations by quarter for  $PM_{2.5}$ .

As an example, here is Glacier National Park which is right close to me for  $PM_{2.5}$ . It shows the percentile distribution of the 24-hour average by year. Notice there is a very good consistency, and then you have your episodes that are occurring, sometimes fires, sometimes other things, but you have a fingerprint of natural variability, in some cases, at the high end of the distribution so that you can, through the approved network and if you pick your sites right, get a pretty good idea of what might be going on concerning the variability of  $PM_{2.5}$  at some very clean western sites.

Conclusions for PM background: A large degree of 24-hour PM variability exists. In its June, 2001 draft, the EPA staff paper stated that for case risk estimates, the Agency would select the midpoint of the appropriate ranges of annual average estimates for PM background levels presented

in the CD.

The use of an annual average smooths the episodic natural PM events such that it is impossible to take into consideration these events when accumulating the daily PM values in a risk analysis, and the variability of 24-hour average for PM for clean western sites should be presented in Chapter 3. It is not appropriate that the staff paper use annual average estimates for PM background levels.

Finally, Chapter 5, the use of the daily average exposure is a controversial approach. Evidence that daily community average exposure has any health significance beyond its high correlation is lacking. Because of the existence of spatial variation and the growing evidence for a threshold, i.e., a nonlinear response, the use of a daily average exposure is inappropriate, and the present language and discussion relating to daily average exposure should be modified as proposed in our detailed comments.

Finally, in conclusion regarding Chapter 5, Dr. Wayne Ott had identified something like 100 errors that were within Chapter 5. Many of them were simple editorial errors that dealt with equations being wrong. Had nothing to do with that is your perception or my perception. Only 5 percent of those changes were made. So, once again, please focus on Chapter 5 regarding some of these errors, and Wayne will be a lot happier.

Thank you.

**MR. FLAAK:** Thank you. So, our next

speaker...you can't keep track without a score sheet here...

**DR. WOLFF:** I just want to make a comment. When we had our teleconference back in February, I was left with the impression that the Agency was going to come up with a distribution of background concentrations to use instead of the single numbers. Any progress been made on that?

**DR. HOPKE:** That was in the risk assessment, yeah.

**SPEAKER:** Yeah, we have considered that in the risk assessment draft that has been distributed and possibly can be talked about later.

**DR. HOPKE:** Right, that's the other...that is in the other half of the equation.

**DR. WOLFF:** But it should be included in the Criteria Document.

**DR. HOPKE:** Right, right. Mort?

**DR. LIPPMANN:** Looking over the previously submitted comments, Dr. Lefohn, I don't see anything about his claim that there is growing evidence for non-threshold, and I would like to receive whatever evidence you think he has showing that growing evidence.

**DR. LEFOHN:** It is in our comments.

**DR. HOPKE:** Okay.

**MR. FLAAK:** So, now, we are going to hear

from Dr. Rebecca Klemm.

**DR. KLEMM:** Hello. I am going to make this, actually, very short. I won't be ten minutes.

I have a couple things that I would like to leave with you as thoughts. What I want to do is update you a little bit on an article that is included in the document that I am the primary author on, Klemm and Mason, where we give the interim results of the Atlanta area study. Mitch Klein is on that same group of people looking at morbidity, and we look at mortality.

First comment is that our article is listed as using GAM which is not true. We did specify natural splines in the document, and it was inappropriate listed there. That is just a correction that needs to be made and, hopefully, it can be.

We have been looking at, for two years now...there were some comments made before that paper was published that because it was only one year, the data probably wasn't very meaningful, and we were asked to put such a statement in the paper before it was published. We did. We now have two years of data, and the evidence is very similar, and that is partly what I want to update you with.

I don't think it is the amount of time. I do think that, in fact, there is a lot of complexity that we still don't understand, and I think you have heard a lot of that today.

I want to tell you a couple of specific things that we have found there, because we do have some interesting

specialized kind of data on mortality that is not available in most of the mortality studies. We did collect actual death certificates. They were redacted by the counties. We got them contemporaneously. We had them before they would be on a national data set. We also, actually, have information about the decedent that far exceeds the amount that you would have on a national NCHS data on mortality.

We have been looking at issues about where the people actually died, whether they were in institutions, and various kinds of characterizations of their health before they died. That is just, I think, things that should be looked at later on and known about.

We do have two years of data which is really just an added one year of data from what was in the interim data in the JAMA article. It spans from August 1st, '98 to July 31st, 2000. What we have found and also in this area data base, I think, that is particularly interesting to note is that we have addressed the lost pollutants. There are more to address, there is further to go, but the interaction of the various pollutants is where our effort has been.

So, there is a list of 16, 17, including  $PM_{10}$ , different air quality indicators that we have actually looked at, and by looking at the relationships with PM and these various different pollutants, we have further things that we think raise questions for further investigation.

With the results we have, single-pollutant model results...again, this is natural splines. You have heard the

difference. We have looked at all kinds of different arrangements. We have looked at GAM to see what that would do. It does, in fact, as most people have said today, raise the coefficients and increase the T values slightly across everything. We have looked at smoothing splines, we have looked at B-splines, we have looked at different knots in the natural splines, and the knots do not have much effect. But what I am presenting here is natural splines over two years with seasonal knots.

The only pollutant on that list that actually is associated with the natural spline generated result with a T value greater than 2, to make it simplified, is OHC, and that has a value of 2.45. That is different from the first year when, in fact, it was negative. So, it actually has come up over two years as positive and the only one with a value over 2.

When we look only at the people who are at least 65 years of age when, in fact, they die which is about 66 percent of the decedents from non-accidental deaths in Atlanta over this time period, we find three. We do find PM 2.5, CO, and OHC with the different values, and then we look at them together, each one of the pairs, to see the effect that they had together which, I think, is where a lot more has to be looked at to understand. I am a statistician. I understand all the complexities of the problems of multicollinearity. We all are learning a lot about that. I am

glad to see that discussed.

But at least to try and tease, beginning to tease out these relationships, we have looked at them in pairs to see what, in fact, happens as they are introduced and then try to understand what might be going on.

When PM is alone, the value on the previous slide which is the same as the T value of 2.52. With CO and with OHC, the coefficient is below 2, the T value associated with the coefficient is below 2. CO alone ended up with a T value of 2.28 when it is alone. When both CO and PM<sub>2.5</sub> are in the same model, it also is below 2, so both of them are reduced in terms of their statistical significance. And with OHC, it is also below 2.

But it doesn't mean, as we can see on the next one, that just because you have multiple pollutants, they also count in terms of their T statistical values or by associated significance, however you want to say it.

OHC alone, a T value of 2.66. With PM<sub>2.5</sub>, the value is actually raised to 3.29.

Now, I have no explanation as to why that is. I am just present you that these things do happen in all sorts of ways, and digging further into understanding some of these things is an important aspect to pursue further on.

With CO, OHC drops slightly but stays fairly close. It is at 2.47.

So, one of the things, then, we are going to be

looking at later and I would propose as being useful for other people to do also is to look at the various components of OHC to see if there is anything in particular...we have daily data now on that that we will be looking at to see if we can understand better which of these potential components of OHC may be, in fact, the thing that is reacting with PM more and bringing up the significance overall.

One of the...tried sort of OH-OHC which we had not seen before but felt now it was worth trying to understand what is going on, and, of course, OHC does spike much more than PM. Here is a chart of the OHC daily values versus the PM, and you see we have the high spike. The correlation between...the simple correlation between the two series is very low, but when, in fact, we remove some of the very high spikes of OHC, the correlation, of course, is much higher. It goes from about...up to about 0.3 and starts at about 0.01.

Simple correlations tell you only part of any story, and I just want to leave you with that message in mind to everyone. Most of you, I am sure, know that.

So, in conclusion, I just want to reiterate some of the things we have heard today and, actually, have been said for many years, but I think some of them are being listened to in a larger context today about some of these modeling concerns and issues that, in fact, have been mentioned, but today, we have heard a lot of very good illustrations of some of the complexities of the computational method that people



have been using which, in many ways, were not very well understood at the time they were first used.

We do see, at least in Atlanta...and I have seen this in other cities, but I am talking about Atlanta here particularly...the association of  $PM_{2.5}$  and mortality in decedents over 65 tends to decrease when, in fact, we have other pollutants in the model. Now, that doesn't mean that it should.

That doesn't mean that it isn't important, but I think it is a fact we have to keep in mind and, later on, understand what is going on there.

Correlation, simple correlation, or even adjusted for extreme values, as we have seen some effects on that today, does not provide sufficient understanding. We need much more than that within groups, subgroups, and times and locations.

And, of course, we don't know what the correct variables are, I think, looking at the components, and I am very happy to be able to be working with data where we have a lot of AQI components available so it isn't just a collection together that we are looking at to get a sense of where these things are coming from, if anywhere in particular.

And I would quote from other people, continue to look at more multi-pollutant models. Even knowing the calculation complexity inherent in that, it gives us a better understanding and consideration. It is what we need to know.

Thank you.

**DR. WHITE:** Rebecca, have you looked at the connection between OHC and carbon at all, particle carbon?

**DR. KLEMM:** We have looked at all the combinations, and I only presented these three, because they are the ones where the bounds were at T values greater than 2.

OHC does not change with a lot of...I can't tell you the exact, although I do have it, but I don't remember without...

**DR. WHITE:** I was wondering about in the air, not through the health effects, but just how does OC vary with OHC?

**DR. KLEMM:** I certainly have that. I certainly have all the correlations, but it is only the temporal correlations I have here. But I do have it, and I can give that to you.

Yes?

**DR. MILLER:** Being one of the few studies that has death certificates, to what extent have you seen cardiac arrest with a subsequent statement about ARMS as secondary to that? I have long-gone concern relative to ARMS in 300,000, more than 300,000 people dying, and I haven't been able to get a straight answer yet still on a number of aspects of where that may be a confounder.

**DR. KLEMM:** Dr. Miller, I can look that up at my office. I don't have that with me, but I would be very happy to pass that on to you. We do have all of the coded text versions of the death certificates not only as they have been coded by a nosologist who is overseeing the cause that you find on the series but also the text, and we can actually answer that question, and it has never been posed specifically, but I appreciate it. We are in a position where questions like that can be answered, can be asked and answered. It is just that we haven't been.

Anybody else?

**DR. HOPKE:** Okay, thank you. Our next speaker, then, is Dr. Fred Lipfert.

**DR. LIPFERT:** Good afternoon. We have heard a lot about numbers today, numbers, numbers, numbers. I am going to talk a little differently about concepts, but I'll give you a few numbers, too.

These are some of the problems that I saw in reading this Criteria Document, questions I felt were important but were not addressed, and I think they ought to be.

First of all, Les told us this morning what was in the Criteria Document, but he didn't tell us how it got there. He didn't talk about the process, and I am very concerned about the process. This is the third version of this Criteria Document, and it still has the same problems of ad

hoc selection of studies, no clear method spelled out about how they were chosen and why they were chosen. It is not a systematic review as defined in the medical literature; it is an anecdotal account, and I don't think that is what the framers of the Clean Air Act had in mind.

In my longer handout which has just arrived, I found 180 citations that I know about that are not in there. Now, I am going to make a positive...it is one thing to stand up here and complain, but I am going to give a positive suggestion which I hope Les will seriously consider.

That is I think the system needs to be changed. Instead of doing PM, CO, or ozone or whatever one at a time when your contractor knows very well that he is supposed to find the studies that meet your requirements, EPA should be doing all pollutants all the time. They need to have the staff set up to do this. It is not that hard. I have been doing this kind of review work for about five years now for some clients. I have 400 citations, one or two-page reviews, and I am not working full time at it.

So, please take this kind of suggestion in mind. We have had enough. You know, three times and it still have the same problems. Ron told you what was done with the veterans study. In our Philadelphia time-series study, we have a table that lists 75 different pollutant results, PM pollutant results. The Criteria Document picks one of them. It is the highest sulfate one, of course.

Let's go on. On time-series studies...have you ever

thought about this...people whom we think died from air pollution are going along day in and day out fat, dumb, and happy, and all of a sudden, boom, they are dead from a pollutant that they have experienced many, many times before at the same levels. You know, study after study finds that it is not the high values, it is the mid-range values that are statistically significant.

Why should this be? Well, it is clear that there is a missing element in the model, and that missing element is the health status of the individual which we don't know, because it is an ecological study, but there are three studies that address this question. Two of them are harvesting studies which specifically analyze the dynamics of the frail population. The third one explains...is an animal tox study by Bob Hankersly that was in the EHP in January that specifically looks at how animals behave near the end of life when they start losing homeostasis.

You put these two things together, and you have an answer. These people are on the edge. Something pushes them over the brink.

Now, this doesn't tell you which pollutant to control, but it gives you a handle on the mechanisms, and I am really disappointed that we don't have some discussion of at least the ideas involved in how air pollution can kill people who were otherwise previously healthy.

Going on down here, we have heard a lot about the colinearity question, and here it is again. I haven't really

see this discussed anywhere, and the problem is there isn't a good answer to this question. We know that the single-pollutant models overestimate the effect of that pollutant. The multiple-pollutant models can give you the wrong answer if there is a difference in measurement error, if one pollutant is spiking and the other one is not.

What are we going to do? I don't know, but somebody should think about it.

Now, let's shift into long-term studies. I know I am running out of time, so I am going to just go right down here to the numbers. We don't really know the mechanisms for long-term studies. We don't know whether they are really acute summed over time or chronic, but there is one piece of information that is right out there lurking for you, and you don't have it.

I have heard today several people say well, when we talk about the cohort studies, of course, we mean the Six Cities and ACS. Hey, I've got news for you. There are a lot of others. I handed you out one today. I hope you will look at it.

There are five, as I make it, and what I have done in making this table was just take a simple average of all five where it applied to, by gender...these are all causes of death. And, of course, the reason males are not as high as females is because the veterans study is all males.

But just look at these numbers now. This number is almost exactly the same size as the time-series number. The

big numbers here are for ozone, but not everybody has looked at ozone. The American Cancer Society did not look at daily max ozone. Only Oshwag and the veterans study and Michael Oshiwa's study have.

These numbers are essentially all the same. So, it is not really clear, when we look at multiple studies, which is the bad actor here. I was very pleased to hear the emphasis this morning on stumbling across time-series studies in order to eliminate the problem of the error of each study. Well, here is another example. Thank you very much for setting that up for me.

I would urge that, instead of the Criteria Document emphasizing one study that they happen to like, let's give them all a fair shake and look at them together.

I think I am going to...how much time do I have?

**MR. FLAAK:** Four minutes.

**DR. LIPFERT:** Oh, okay, thank you.

Children's health, a big issue. The Criteria Document talks a lot about the study by Tracy Woodruff on infant mortality. She finds that PM<sub>10</sub> is responsible for SIDS, sudden infant death syndrome, which I think is just totally irresponsible.

We looked at that study and replicated, and, of course, the first step in a replication is to take your data and the other person's method and see if you can get the other person's answers, and we did. That is what is in the Criteria

Document. Lipfert replicates Woodruff, big news.

The big news is that sulfate is extremely protective of SIDS. The reason for that is because SIDS is high in the West and low in the East. In my opinion, it has absolutely nothing to do with air pollution, but if believe Woodruff, then you ought to be fumigating your child's bedroom with sulfate aerosol. Do I have to say anymore?

Now, the next point has to do with intervention studies which I think are very important, because neither the time-series studies nor the cohort studies are going to tell you whether things will get better. The intervention studies in there don't really deal with PM<sub>2.5</sub>.

Let me go on. You know, you have had NMMAPS until you are blue in the face. Well, sorry about that. Here it is again. I am going to have to go fast. I think what I will do is jump down here. I want to show you some plots.

Oakland and New York are the two significant points. Well, we thought about this a little bit. In fact, it was Ron Wyzga who pointed this out to me. In 1991, Oakland had a big problem with fires. We don't know what the air quality was then, because AIRS doesn't give you the daily value anymore, but I would suggest that one ought to test whether 1991 is a significant year for Oakland.

And as far as New York and the Northeast goes, 1988 was a severe heat wave here. The mean PM<sub>10</sub> in New York City for 1988 is about 50 percent higher than any other year. So,



we saw an excess in the Northeast. That is where this heat wave was. I don't think the Hopkins model takes long duration heat waves into account. That's a hypothesis.

Now, just a couple more, the question of how does the T value change. Well, with due respect to Suresh, you showed the same plot turned around the other way. I fit a quadratic to it, because I noticed that curvature in yours, and I find it really interesting. I don't have explanation. It gets turned around the other way from what Suresh had.

The negative values are not affected. The positive values are affected a lot, to the point that if you use this relationship as a calibration for studies that use the old GAM but you don't have information for the new GAM, you would say you would have to have a probability of 0.001 in the old values to be significant in the new ones. That is a big change.

Okay, I have other plots if you want to see them.

Thank you.

**DR. HOPKE:** Thank you, Fred. Now, our next speaker, then, is Anne Smith from Charles River Associates.

**MS. SMITH:** Thank you. I am speaking today on behalf of Edison Electric Institute, and I also want to introduce my colleague who collaborated closely with me on this, Dr. Tim Savage, who has a background in econometrics

and nonparametric studies in particular.

The key points that I want to make, and then I will go through them in detail, but, first of all, as we know from today's meeting already, the problem that has been found in NMMAPS is widespread throughout the literature that is being cited in the PM CD, but more importantly, what we would like to highlight is that we think this may just be a harbinger of a much broader class of problems that are related to numeric accuracy in the types of statistical techniques that are being used.

With that in mind, emphasis on numerically intensive and new or n-dimensional statistical methods do present a serious concern for numeric accuracy, and I am going to describe why. More generally, what I think is needed in the PM CD is that the...well, actually, in the epidemiology literature generally is more focus on getting some insights about what is going on underlying the many different studies that we are getting and relying more on transparent and traditional methods as well as the ones...the sophisticated ones to make sure we get some insight about what might really be the effects so that we can protect public health best.

From the point of view of risk analysis, I would like to just say that it would be misguided to try to perform a risk analysis that uses just current results out of the current literature. There is far more uncertainty than is being represented in any statistical error even if you look

at a meta analysis across all of them.

The first point, then, there are many studies...actually, Lester Grant's handout this morning listed many studies which appear to have the same problem. This is our list, and it is really just the studies that appear to have a problem from Tables 9.14 and 9.17 of the Criteria Document. We feel that these are the highest priority for checking and reviewing, making sure that the problems there get resolved before any of those studies are relied on again.

We heard a lot about bias today. Working without the benefit of all the information that I have heard this morning, our position was that there is evidence that the bias created by that GAM problem was in the upward direction. We have seen it directly in the reanalysis of the NMMAPS results, and taking into account the comments that have been made by Dr. Burnett prior to my presentation, I also wanted to give you an example where we may be able to see the same effect going on of an upward bias in the apparent effect when using a nonparametric approach as opposed to a parametric approach.

I'll come to some examples, but the key point I want to make here is that the effect of this convergence problem is essentially as if we haven't really done the controlling that we were intending to do in these studies. So, the coefficient or the relative risk for PM will be biased due to lack of control for important covariants to the extent that the convergence kind of ended prematurely.

In a sense, that means that the recent literature, since 1996 where most of the effects have appeared, where the PM method has appeared, is not necessarily any more controlled or any better controlled than the earlier literature from the previous Criteria Document.

Now, just quickly looking at what went on in the Krewski version of the ACS analysis versus the recent Pope, et al analysis, the first thing I show here is from the Krewski report which uses parametric controlling for spatial autocorrelation. Without the controls for spatial autocorrelation, the Krewski replication gets pretty much the same result as the old Pope one. That is the replication.

Now, in Krewski, et al, when they add a nonparametric...I mean, sorry, a parametric method for controlling for spatial autocorrelation and remove that spatial autocorrelation, the all across mortality effect falls by about 75 percent. The significance falls dramatically as well.

That is an example of a fully controlled regression on the ACS data set where we have not got the spatial autocorrelation.

Now, in Pope, et al, 2002, the more recent one, we don't see as big of a drop. Now, this is the one where the nonparametric method has been used.

I find it interesting, given that there was apparently no significant autocorrelation problem, that, in fact, we do still see the drop. I think that begs the

question of what is going on, and if we look at the cardiopulmonary mortality, it is even more pronounced in the original data. We saw it drop to non-significance, not quite as large a percent drop, but it still occurred when we controlled for spatial autocorrelation, and in the Pope, 2002, no such response happens.

I think this is really the case when people quote Pope, 2002 and say there has been no change in the...no effect on the result as a result of the spatial autocorrelation fix. That does show up in the cardiopulmonary case, but I think it raises a whole other question here, whether or not there is any spatial autocorrelation in this data, is why did it go away? Basically, the same data, the same people, just a little, you know, twice as many years of data.

Why would a strong spatial autocorrelation disappear with eight more years of data? I think that poses an interesting question that ought to be explored more. I don't know any reason, and I don't have any hypotheses, but it suggests that maybe there is something non-stationary in whatever effect is underlying the ACS data. It might have something to do with exposure or measurements of exposure.

But the more important issue I want to get to, the more general one, is that we really need to thoroughly explore numeric accuracy in general. This is because computers can produce very different results for the very same set of data, and I can depend on the sequence of

arithmetic steps that are programmed into the code. Different codes will have programmed the same statistical technique different ways. The sequence in which the arithmetic steps are done by the programmer can end up giving you different results out of the computer, because the computer programming logic in mathematics works differently than true arithmetic.

Also, it can depend on the format in which the data are entered as well as the precision of the computer itself. None of this has anything to do with bugs. It is just inaccuracy that is associated with computers.

These inaccuracies will occur in statistical software far more often than most people think, and this has been brought out in the econometrics literature just in the last couple years in a very useful paper that I want to give you a couple quotes from by McCullough and Viard from 1999. It really emphasizes this. But the key thing is that PM epidemiology is very programmed for this kind of numeric inaccuracy, and that is why we thought it important to raise it.

How is it fertile ground? First of all, the bullets here represent the situation that McCullough and Viard bring out as important for engendering inaccuracy numerical. Small values being estimated, that is clearly what we are dealing with in a fraction of a percent on a small percent risk of mortality.

Frequent use of ratios, ratios are more susceptible

to problems than subtraction and addition, as the proportion of hazards are ratios. So, the very thing that we are estimating is a ratio.

Many, many iterations of numeric intensity in the calculations, and this points directly at almost all the methods that are being used, but maximum likelihood information as well as nonparametric methods. This goes well beyond fixing the convergence criteria in the GAM. It also goes well beyond the GAMs. It has nothing to do with convergence criteria at all.

And use of random number generators shown to possibly get people into problems. The MCMC algorithm from the Bayes information involves random number generators.

Unconventional statistical techniques, by this I mean relatively new ones that are just entering into practice are also important. They have the least testing.

Just a few excerpts from their paper. The point...I'll just end on this final point: The user should always have some idea of the software's precision of range and whether his combination of algorithm and data will exhaust these limits.

The recommendation that we have is EPA should really engage in a thorough testing and bench marking of the software that is being used, and that means bench marking and testing with the kind of data that are being used in the PM epidemiological studies. As McCullough and Viard say, to fail to test that software represents the triumph of hope

over experience and is an invitation to disaster. They give quite a few examples numerically that explain why.

Now, just to turn for a second to the Bayesian pooling, we have seen this graph before. I am repeating it for a second. The top bar is the non-Bayesian manipulated MLA controls with their error bars, and you can see they kind of tend around zero, and we can just see the miraculous transformation that is really quite counterintuitive even to people who are familiar with Bayes and Bayes 4, how it works. It is not well documented, and this really does demand explanation.

So, we feel that these results not only require more explanation, but the numerical accuracy of the software requires close inspection, and we feel that the more important point is pooling is premature anyway even if you were to use more intuitive or conventional methods. The real question here is why is every...the PM effects so different from city to city. That is what I think all the effort should be put into.

Thank you.

**DR. HOPKE:** Any questions? Warren?

**DR. WHITE:** What journal is the article in?

**MS. SMITH:** Is it JAMA, I think? I have the reference for you.

**DR. HOPKE:** Okay, the next one will be a



tag team. We are going to get to start with Dr. Jay Turim, and then Dr. John Richards will take over after five minutes.

**MR. FLAAK:** Dr. Moolgavkar would have been next, but he has graciously said he could wait till tomorrow, so we'll schedule him for tomorrow.

**DR. TURIM:** I appreciate that. Well, good afternoon. I am Jay Turim, and these comments are being submitted on behalf of the National Mining Association, National Stone, Sand and Gravel Association, Industrial Minerals Association of North America. Will DiCalco helped me prepare the comments, and I will be speaking quickly for five minutes, and then my colleague, John Richards, will be talking five more minutes. So, if you will, let me know when my time is up.

**DR. HOPKE:** Sure.

**DR. TURIM:** The issue that I will be addressing is the suitability of the data presented in the third CD to the assessment of the exposure-effect association between current levels of the PM<sub>2.5+</sub> fraction which we call PM<sub>C</sub> for health effects and the attendant question of whether the CD supports data that can be used to prepare a quantitative risk assessment for individuals exposed to PM<sub>C</sub>. I want to emphasize that our comments are directed at only the coarse fraction. We make no comments...we are not questioning the suitability of the Criteria Document for PM<sub>10</sub>

or PM<sub>2.5</sub>. We are restricting ourselves to the coarse fraction only.

It was about a year ago when I addressed the same body, and my comments then were about the limited number of studies that were available in the second Criteria Document dealing with PM<sub>C</sub> and the limitations of those studies. We were gratified that CASAC, in their letter to EPA, acknowledged our major concern with the absence of critical information on PM<sub>C</sub>, and we think that statement should be put into the new letter that CASAC sends to EPA, because, although some studies have been added pertaining to PM<sub>C</sub>, we have found nothing new to alter the opinion, namely, that the document is not sufficient to support compellingly a dose-response relationship.

The current studies are not adequate to demonstrate an exposure-effect relationship between PM<sub>C</sub> and mortality. Of the 12 new studies published since 1996 in which exposure to PM<sub>C</sub> has been evaluated, in only three, Phoenix, Santiago, and Mexico City, have there been any statistical associations shown, and of those, one of them, the Phoenix study, has been marginally significant.

Even those studies, the CD points out, the ones that have been statistically significant are pointed out by the CD to have problems. For example, in the Phoenix study, the Criteria Document states that biogenic processes may

contribute more to observed  $PM_C$  effects than other particles, and the CD states that this entanglement of potential contributions of biogenically derived organic particle components from those of crustal particles in Mexico City and Santiago pose challenges.

The situation with respect to morbidity is not much better, we don't think. It is true that since 1996, a number of morbidity studies have been added to the...which have allowed investigation of the effects of PM on morbidity outcomes. However, it has been very, very difficult to tease out of that information contributions made by individual components of the air pollution mix, especially  $PM_C$ , and we think that the current evidence is insufficient to associate coarse particle,  $PM_{10-2.5}$ , with morbidity effects.

There are many different reasons why it is difficult to associate  $PM_{10-2.5}$  with health effects, and I will just talk about one or two of them. These are covered in the Criteria Document. Over lunch, I was counting the number of citations in the handout that Dr. Grant gave this morning, and there were 400 epidemiological studies, approximately, of which about 5 percent related to  $PM_{10-2.5}$ .

Well, one of the reasons it is difficult is because there is a variability in particle composition among different regions and with seasons that have not been adequately accounted for in many of the studies. I won't go through the reasons, but they are in your handout.

The exposure data used in the  $PM_C$  studies are highly suspect, and my colleague, John, will be talking more about exposure monitoring.

So, I will just conclude with this first half of the presentation by saying that the available studies are inadequate to associate current  $PM_C$  levels with human health effects. We have highlighted five or six reasons the number of  $PM_C$  studies that have statistical significance is small. The available studies demonstrate, at best, only a weak or equivocal association between  $PM_C$  and both mortality and morbidity. The studies are confounded by the presence of other pollutants. The exposure data on which the studies rely are highly suspect. The studies don't accurately reflect the relative prevalence of different particle effects.

These are among the reasons why it is very difficult to associate human health effects with  $PM_C$  exposure.

Thank you. Thank you for the time.

**DR. HOPKE:** That's fine.

**DR. RICHARDS:** Good afternoon. I would like to make just a few more comments to Jay's comments. Again, all my comments are directed to coarse material only.

There are three major concerns I am going to very briefly discuss. One is that there is very limited presently available coarse data or  $PM_{10-2.5}$  data available. There is

very little speciation data, and also, emissions data are also very sparse in the document.

Number two, there is insufficient data concerning the spatial and temporal variability of the coarse particulate matter. And number three, of course, has been discussed a lot. There is very little coarse data available for use in epidemiological studies, and I think that is an unfortunate limit.

One of our major comments is that there is very little data in section 3.2 and appendix 3A concerning the coarse data. The data that we do have available is primarily in the form of difference data, in other words, a  $PM_{10}$  measurement minus a  $PM_{2.5}$  measurement. That is not a very accurate way to assess coarse particulate material.

Number two, even the difference data effect that we have is very limited and does not give us a good basis for evaluating exposure in various urban and rural areas throughout the U.S. So, we have two major concerns about the quality of the coarse data.

There are also brief discussions in the Criteria Document concerning emissions, and that data is very limited with regard to the coarse data. In fact, again, there has been a stack test method out for four or five years on measuring coarse data directly. Unfortunately, almost none of the measurements that have been made or the papers that have been discussed have been referenced or used in the

Criteria Document. So, there is additional data that would help evaluate emissions of the coarse data.

Again, some of that emissions data would be very useful to evaluate the relative importance of natural versus anthropogenic sources of the coarse material which is a particularly important issue with regard to coarse. I don't think you will see it as much in the fine area.

Very briefly on the natural sources, just a few points that I think could be expanded upon in revisions to the Criteria Document, is there is just a very limited discussion of sea salt and its distribution around coastal areas. There could be more on wind erosion in rural area climates, and that would be particularly important in the West. Additional information on forest fires, including some mineral particulates, not just the organics from forest fires, and also, of coarse, something that is discussed in some detail is the global transport of the crustal particulate matter, and some of that, of coarse, is in the coarse fraction.

Since Jay touched on this, I will go fairly quickly across this. Again, the coarse data is quite different than PM<sub>10</sub> data or PM<sub>2.5</sub>. I think that is discussed in the Criteria Document, but, again, some of the reasons that the PM<sub>10</sub> data are not a good indicator include the well-discussed spatial variations in the coarse particulate, the differences in the formation mechanisms, and the fact in the East, we are

about 75 percent fine material in  $PM_{10}$ . So, the coarse material is a small fraction of the  $PM_{10}$ .

Two recommendations to conclude. One thing I think would be very helpful in the process would be the development of a reference method to directly measure coarse particulate matter. I think that should be given some priority. Then, with that monitoring technique, obviously, there is a need to compile considerable information concerning spatial and temporal variability of the coarse material, its constituents, and also the characteristics of that material, not just the concentration.

So, those are our main recommendations. Thank you.

**DR. HOPKE:** Clarification. You are suggesting that there are data on emissions that aren't in the CD, but you are not suggesting that there are ambient concentration data that exist in the literature that aren't in the CD?

**DR. RICHARDS:** The question was, are there emissions data available that aren't in the CD, and that is basically true. There are some published papers that have coarse primary emission data. I think it would be helpful to look at the relevant importance of primary emissions versus natural emissions. I would recommend that be got into.

I was not referring to the ambient data. All the data that is available, I think, is in the paper, but our comment, again, is there is not much of that to be had.

**DR. HOPKE:** All right, thank you.

Okay, our next speaker, then, is going to be Dr. Ferdinand Vendetti.

**DR. VENDETTI:** Thank you. My name is Ferd Vendetti. I am chairman of the Department of Medicine at Albany Medical College. I am a cardiologist and a cardiac electro physiologist, so I am really feeling like an odd duck in this room today. A lot of energy around statistics which I am not going to go anywhere near.

I am here as a consultant for EMA, and I am going to talk to you a little bit about some of the cardiac studies that are discussed in Chapter 7 and Chapter 8. I think my background might provide a slightly different perspective and one that the committee has not heard before with regard to this data.

I have had an opportunity to go through all of the studies in a fair amount of detail that are reviewed in the document, and I have a fair amount of concern about a number of issues, not the least of which is some of the methodological flaws from the perspective of a physician, of an electro physiologist who has had a lot of experience in many of the areas that are reviewed in those two chapters.

Let me start with just four of the research studies that are reviewed, and I would drop first to the bottom, to the Peters study, which is looking at ICD therapies as a surrogate for sudden death. I implant ICDs. That is an acronym for implantable cardiac defibrilators. It rescues



people from sudden cardiac death episodes.

As an individual actively involved in caring for such patients, I can tell you that, as a surrogate for sudden death, it is wanting. As a matter of fact, the E.P. community has gone so far as to have a multi-center controlled trial to actually demonstrate the efficacy of this therapy, because, as a surrogate, a device shock is very inadequate in terms of telling us the patient is going to die.

Nonetheless, this particular study used that as a surrogate for sudden cardiac death in patients who were exposed to various levels of air pollution, and I think that, as a result of the potential for inappropriate shocks which are not related to arrhythmia, shocks for device malfunction, shocks for rhythms other than life-threatening arrhythmias, that the conclusions of that particular study are somewhat suspect and need to be further either modified and looked at, or looking at actual studies of sudden cardiac death would be much more appropriate.

In addition, there are methodological flaws with regards to the use of heart rate variability. Heart rate variability I could spend an hour up here trying to explain to you. Basically, it is a measure of the beat-to-beat change in heart rate. This is an indication of autonomic modulation of heart rate. In many prospective studies performed in patients with cardiac disease, it has been demonstrated to predict outcome.

Heart rate variability is very variable, and as a methodological issue, these are several different of heart rate variability. They are done in five-minute blocks of time, and one can see, if you look at the middle graph which is high frequency power which is an indication of vagal enervation or vagal changes in the heart, that there is enormous variability. This is during sleep, this is during awake hours. Obviously, during sleep, there is high vagal tone, and we see that reflected.

If we take an hour or two, however, when someone is up and about, you can see that there is still a several-fold variability, three to four-fold variability, in the course of an hour in five-minute intervals.

Many of the studies cited in this draft document use very small sampling, five- minute samples, six-minute samples. One could imagine that if that sample was at this point versus this point, there might be tremendous variability introduced into the analysis simply based on what is being measured.

In my discipline, when we look at heart rate variability, we look at, typically, a mean of 24 hours to get a true reflection of the autonomic modulation of heart rate.

The other major issue with regards to heart rate variability really goes to what people or what the investigators think heart rate variability implies, more specifically, that acute changes in heart rate variability might actually be associated with acutely with an increased

risk of sudden death or a cardiac event. In point of fact, that is not the case.

This is data from a study where, fortuitously, or, for the patients, not so fortuitously, 24-hour whole term monitoring was being performed when there was an episode of sudden death. In those 24 patients, the whole term monitoring was then analyzed for heart rate variability changes. What you have here illustrated are six of the patients who succumbed when they had their whole term monitors on.

If you just look at these six graphs, that is a very simple time domain measure of heart rate variability. The arrows are the point in time at which sudden death occurs. This particular patient was fortunate to be resuscitated.

There is no pattern over time. There is no acute change that happens within five minutes, ten minutes, an hour of that sudden death episode. The investigators actually looked at the first hour of recording and compared that to the hour prior to death. There was no statistically significant change. They then compared this cohort of 24 patients to 19 patients who were age matched and matched for morbidities, and, again, there was no difference in their heart rate variability parameters.

Putting aside the issue, the methodologic issue, as I researched this topic, it was clear that there were a number of studies not reviewed in the document. One in

particular, Checkaway, a study that was done in the Pacific Northwest, there was a very interesting study that was funded by HEI. I was kind of surprised it wasn't juxtaposed to the Peters study.

This is a study of 362 sudden cardiac death or cardiac arrest victims or subjects, and it is a case crossover study where they looked at PM content on the day of their arrest versus a matched day within a relatively short window and found no effect of PM concentration on the incidence of sudden cardiac death.

In addition, the study by Brewer actually looked at a group of patients with prime obstructive pulmonary disease, monitored 24-hour heart rate variability at the time of personal...I am sorry, I am not familiar with the terminology...individual monitoring device, actually have their PM exposure, and found, again, no significant correlation between heart rate variability changes and PM exposure. That particular study evaluated individuals on seven separate occasions yet did not discover any association between heart rate variability and PM.

There are a number of studies that have been discussed that demonstrate no effect on heart rate variability. There is even one study which, I believe, has been around for a while and talked about that actually demonstrates a beneficial effect. Heart rate variability goes up. T wave alternating, which is another measure of that phenomenon, decreases. Both are a sign of improved

cardiovascular fitness, if you will, as opposed to worsening cardiovascular fitness.

And, finally, in several of the studies, within their own data, there are conflicting results. Parameters, two different parameters that should be very highly correlated in disagreement, heart rate changes in opposite direction to heart rate variability which, again, usually are very tightly correlated.

So, I would just conclude by urging the committee to look at this data very carefully and very critically before drawing conclusions about cardiovascular effects that might play a role in enhanced mortality.

Thank you.

**DR. HOPKE:** Any quick questions?

(No response.)

**DR. HOPKE:** Okay, our next speaker, then, is going to be Dr. Jon Heuss, Air Improvement Resource.

**DR. HEUSS:** Thank you. I reviewed the draft CD for General Motors. There are several important issues that we feel are ignored or downplayed in this third draft, and these include issues that were raised in previous public comments but also issues raised by CASAC and by EPA staff itself. I will have some comments on the epidemiology, dosimetry, and exposure.

We already raised the issue that multiple studies of the same city do not produce the same result and provided

numerous examples in public comment. This is still a major issue related to model selection.

The third draft indicates, and this is quoting, the fundamental issue essentially subsuming all other modeling issues is the selection of an appropriate statistical model. Further, the basic issue is that there are an extremely large number of possible models any one of which may turn out to give the best statistical fit for a given set of data.

The practical ramification of all this is that multiple studies of the same city do not produce the same result as to the pollutant or pollutants which are implicated or even to health endpoints affected. This inconsistency should present a severe impediment to using the data for policy decisions. The third draft downplays this issue.

After the GAM issues are resolved, we think the Agency should confront this issue head on, not sweep it under the rug as is done in the third draft.

In discussing the chronic mortality studies a year ago, EPA staff raised this issue of heterogeneity in the ASC study. Fine PM had a negative association with mortality in the West when the data was aggregated among the four western NMMAPS regions. Now, in the HEI reanalysis, sulfate is also reported to have a negative association in the same western areas.

Even though EPA staff brought this issue up a year ago, the third draft is silent on it. We think it is important to follow up this difference with the updated ACS

data.

It addressed the question of whether the ACS associations are causal and universally applicable. With the analyses available today where you have several positive studies and several negative studies with evidence of heterogeneity in the major positive study, one conclusion we think should be put on the table is that high mortality associations are either not causal or are not universally applicable because of differences in PM composition or long-term cohort effects.

Turning to dosimetry next, when CASAC reviewed the second draft, the committee concluded that the chapter on dosimetry provided extensive discussion of dosimetry models, but, quote, there was no effort to use this knowledge to connect information on exposure, dose, and health effects suggested by tox or epi.

The committee went on to indicate the connections could be greatly improved by providing examples of the magnitude of the positive and retained doses and pointed out this information is critical to setting the stage for evaluating how the tox information might apply to these epi observations.

However, the third draft contains no such examples. We think it should. For example, Snipes, James, and Jarabek use the ICR FEMA dosimetry model. They looked at several different regions, focusing on the AI or alveolar interstitial region, and when they expressed the retained

doses on what I would call an effect related metric, that is, the dose per square centimeter of lung surface per day, it came out extremely low, in the range of 0.18 to 0.025 ng. In addition, Snipes, et al concluded that there is no clear dosimetric motivation for the use of the fine PM fraction in relation to the positive or retained dose.

Last year, Dr. Yaro Vostal presented his extension of the Snipes analysis to CASAC. He showed that the relevant doses for fine PM components are even lower, as you might expect. Things like sulfate or elemental carbon, things like this, are down to a fraction of a picogram per unit of surface.

For things like the toxic metals which are suggested as perhaps the most probable cause of fine particle toxicity, the 24-hour deposits are extremely low, not exceeding tens of femtograms. When I first looked at this a year or so ago, I had to figure out what a femtogram is. It is  $10^{-50}$  gram.

We think examples such as this need to be added to Chapter 6. They provide a needed link between toxicology and epidemiology. The low levels are constrained on any explanations for the mechanism by which PM may cause or aggravate health effects.

Turning to exposure, I pointed out a year ago that it is a major error for EPA to assume that exposure to PM of ambient origin is independent of exposure to PM of indoor origin. The third draft still dismisses confounding by



indoor pollutants by arguing that daily activities are independent of weather by analyses that start by assuming independence and by analyses of the Peking data which is atypical.

Daily activities with emissions that lead to both indoor and outdoor PM are, to a first approximation, independent of weather, but daily changes in weather drive outdoor pollutant concentrations, and they also influence air exchange rates that determine the exposure to indoor pollutant sources. In naturally ventilated buildings where people spend the bulk of their time, weather affects the air exchange through wind-driven and temperature-driven pressure differences.

The air exchange from wind-driven pressure differences is essentially first order wind speed. Temperature differences across the building shell induce a density difference that results in a pressure difference, and the combined pressure differences from these two sources determines the flow through openings and, hence, the air exchange rate.

Plus, reductions in wind speed, with everything else constant, will both increase ambient levels of pollutants and reduce air exchange, thereby increasing exposure to indoor air pollutants.

Now, ambient temperatures and wind speeds vary diurnally. They vary from day to day and seasonally. So, the mix of this wind and temperature-driven ventilation will

vary. Therefore, the degree of potential confounding by indoor sources will also vary.

However, to be straightforward, to simulate this by linking the outdoor dispersion models that EPA has with EPA's indoor model using standard equations from the heating and ventilation community. In addition, EPA has gathered an 18-month data set that could be utilized to analyze this issue.

Now, instead of acknowledging the potential link and examining it, the third draft ignores it, persistently arguing that exposures to indoor and outdoor-generated pollution are independent. While this makes it easy to deal with the issue, dismiss the confounding and dismiss any measurement error implications from it, it is not sound science.

The suggestion that ambient concentrations of gaseous pollutants serve as surrogates of personal exposure to particles rather than confounders is based on a fairly small sample of data. It ignores a substantial body of studies in the literature. This is documented in an appendix to the comments of the Alliance of Automobile Manufacturers' comments on Chapter 5.

For each of the gases, ozone, CO, and NO<sub>2</sub>, there are studies that report correlation of personal exposures with fixed monitor data. These range from, in some cases, not significant to weak to moderate, but they exist, and similar correlations exist for PM. The PM document covers

those. It should also cover these for gases.

Potential confounding by gases cannot be dismissed at this point, and another conclusion you draw from this is that there are significant measurement error issues related to the use of fixed monitor data for all of the criteria pollutants.

There are some minor issues, but we still think important, in Chapter 5. A couple of these have already been mentioned by others. The spatial variability documented in Chapter 3 provides substantially more potential for exposure misclassification than Chapter 5 indicates.

The discussion of measurement error omits an important contribution Dr. Chock presented to you last July demonstrating measurement error of PM in the absence of a threshold.

We would also raise the issue of nitrate volatilization indoors. It is acknowledged in the CD on page 947. There are several papers both in the CD and others that are in the literature, not in the CD, that demonstrate this, and its implications should be discussed.

Finally, biology or bioaerosols need to be included in Chapter 5. In Chapter 9, it is indicated that they are, indeed, among the candidates, so they should be included in Chapter 5, too.

Lastly, we are concerned that the recent time-series studies, once they are resolved, will still leave additional model selection issues. In addition, we have

pointed out problems with interpretation of chronic studies. Importantly, we think dosimetry modeling needs to be added to the chapter and discussed in both 6 and 9, and we think the Agency should reexamine the exposure material that is relevant to understanding the epi after the reanalysis of the affected time-series studies. This would include the NMMAPS studies that look at all the pollutants.

So, we don't think it is appropriate to close on Chapter 6 or Chapter 5 until after that is done.

Finally, confounding by outdoor gases and indoor pollutants, I think, are still important issues, and the CD dismisses these based on either false or incomplete arguments. So, in essence, we think Chapter 9 must be extensively revised to address these issues.

Thank you.

**DR. HOPKE:** Questions?

(No response.)

**DR. HOPKE:** All right. We need to make a couple of decisions. We...

**SPEAKER:** We are going to finish the rest before we finish today?

**DR. HOPKE:** Well, let's decide. I mean, we could do one or two more. Two more would finish off the comments, and we wouldn't have to do it in the morning. I would like, in any case, to start tomorrow 15 minutes earlier to give us a little extra time for discussion. So, would

people be willing to stay here for another 20 minutes and get the last two done? Yes? All right.

**MR. FLAAK:** Let me also identify the dinner arrangements so you are aware of those.

**DR. HOPKE:** Right.

**MR. FLAAK:** Dinner reservations are for 6:15. There is a van outside. We can take you all directly to the restaurant. If there is enough time, we could swing by the hotel first, but if we do the two presentations, there probably wouldn't be.

**DR. HOPKE:** Okay.

**MR. FLAAK:** And there is room for a few additional folks to join us if you wish. So, anyone who wishes to join us, let me know.

**DR. HOPKE:** All right. So, let's get these last two done and be finished with public comments, and then we can start fresh in the morning discussing where we go.

So, our next speaker, then, is going to be Dr. Moolgavkar, and then we will finish up with Dr. Green. Thanks for being flexible.

**DR. MOOLGAVKAR:** I would like to apologize for holding you hostage longer this afternoon than I was supposed to, and I will not do that again, and Dr. Hopke will not allow me to do that again.

I just want to make a few comments about both the

time-series studies and the long-term ACS study. Here is the chronology of PM meta analysis in the last two years. The first original meta analyses using classical techniques estimate an approximately 1 percent increase in mortality for a  $10 \text{ Fg/m}^3$  increase in PM,  $\text{PM}_{10}$ , that is.

Then, the first NMMAPS analyses, hierarchical Bayes using GAM in the first stage, found...estimated about a 0.4 percent increase in mortality. Now, using hierarchical Bayes and GAM in the first stage with the revised convergence criteria decreased that to 0.27 percent increase in mortality for  $10 \text{ Fg/m}^3$  increase in  $\text{PM}_{10}$  from the previous data.

A hierarchical Bayes using natural splines in the first stage using GAM models then resulted in an estimate of 0.22 percent increase in mortality, and as I showed earlier today, if you use hierarchical Bayes, using a flat priors, natural splines in the first stage, with the two outliers removed, the increase in mortality is 0.12 percent per  $10 \text{ Fg/m}^3$  increase in  $\text{PM}_{10}$ . So, you can see the steady decline in these estimates from the first one which, of course, the authors of the first one swore by as well.

So, it is clear that we are chasing smaller and smaller numbers with more and more statistical...more complicated statistical technology.

Let's look at the results of the long-term studies. What is important for us in the ACS II study is the sensitivity analysis conducted by Krewski and colleagues,

and, for some reason, everybody ignores the fact that the single pollutant that was most strongly associated with mortality in that study is sulfur dioxide. It is not  $PM_{2.5}$ , it is not sulfate; it is sulfur dioxide. There are no two ways about it. When sulfur dioxide and one of the particulate metrics is entered jointly into the analysis, the sulfur dioxide coefficient is the one that survives and is significant.

Now, what is the explanation for this finding? I don't know, and I don't think anybody else knows what the explanation is. Clearly, sulfur dioxide is a surrogate for some complex pollution mixture, but so is PM, and that is the way all the results of these studies should be treated. And it is the fact that one should try to find an explanation for this fact. It should not be ignored or swept under the rug.

Now, Pope, et al the 2002 study is often given as an example of...it is often quoted as an update on the HEI analysis, but this study simply does not address the issue of sulfur dioxide at all. It simply punts that issue by not reporting on any joint analysis, though it does say that oxides of sulphur separately are significant in analyses of these data.

There is another problem that has been nagging me for quite a while with the Krewski study, and that is the fairly strong association noted in the 63 studies between cancers other than lung cancer and  $PM_{2.5}$ . Now, this

association is stronger than the association of  $PM_{2.5}$  with lung cancer, and it is highly biologically, it seems to me, implausible, and it suggests to me that despite the care taken in this study to adjust confounding, that confounding, residual confounding, might still be an issue in this study.

Here is my final slide. I think the issue of residual confounding is extremely important in operational studies such as air pollution studies, particularly when the estimated risks or benefits are small. And I think there are a couple of contemporary examples that dramatically illustrate this point.

The first would be the example of beta-carotene and lung cancer and heart disease. Many observational studies show or seem to indicate that beta-carotene was associated with protection against lung cancer and heart disease, and there is a perfectly good biological explanation for this, because beta-carotene...so, there is a perfectly plausible explanation for beta-carotene being protective against heart disease and cancer, because it is an antioxidant, but when a rigorous clinical trial, randomized clinical trial, was done to test this hypothesis, the only way, the only way to provide any kind of residual confounding in the epidemiological study, exactly the opposite was found.

The same thing turned out to be true for hormone replacement therapy and heart disease. This, of course, has hit the news recently, so everybody must be familiar with this example.



And I think it is about time that the epidemiologists recognize the limitations of our discipline, because no amount of fancy statistical technology can overcome the inherent deficiencies of observational data, particularly when we are chasing very small risks.

And I think I must stop there. Thank you.

**DR. HOPKE:** Thank you. Any quick questions for Dr. Moolgavkar?

**DR. BURNETT:** I just want to reassure Suresh that the Health Effects Institute has very kindly sponsored us for another two years to hammer the ACS data, and all and more of those concerns will be addressed as best as possible.

**DR. HOPKE:** Good. All right, our next speaker is Dr. Laura Green.

**DR. GREEN:** Thank you very much. Thank you very much to the committee and to the Agency for allowing me to spend a few moments with you, especially since all you probably want to do is get out of this room and go to the bathroom, but I do appreciate much more generally and broadly the incredibly difficult job that the Agency has and that the committee has in grappling with these issues.

As Professor Moolgavkar said, epidemiology that is observational and that presents weak effects can only go so far. I am a toxicologist and a chemist by training, so I would like to address the toxicologic aspects of this

question of whether and to what extent PM affects public health. In particular, I would like to try to give a few constructive suggestions for the improvement of the Criteria Document, in particular, Chapter 7 on toxicology and Chapter 9 on causation or integrative synthesis.

Briefly, my thesis, of course, is that toxicology is an important part of this puzzle. Why? Because the observational epidemiological studies, by definition, are not experimental, randomized, double blind, or placebo controlled. Now, that can't be, of course. I mean, you can't make people move to, you know, Schenectady and some to Pittsburgh and some to Wichita, so you are stuck with that.

Nonetheless, of course, because the relative risk estimates are small, perhaps getting smaller all the time, perhaps not, there has to be something to sort of pick up the slack. Traditionally and in the minds of still many people, there is a disconnect, a discontinuity, a difference of opinion, if you will, between what the epidemiologic studies suggest and what toxicologic and clinical observations have suggested with respect to moderate level exposures to particulate matter.

So, it seems to me that, fundamentally, what the Criteria Document for particulate matter must do is address, very seriously and explicitly, the \$64,000 question, do current concentrations, things that one could be regulating tomorrow and in the future, do current concentrations of particulate matter, however defined in air in the United

States, cause disease and death in such a way that lowering them...do we believe, and, if so, on what basis, that taking a city's annual average PM level,  $PM_{2.5}$ , let's say, levels down from  $16 \mu g/m^3$  to 14, do we believe that that will improve public health, and, if so, how and in what way?

I would like to just quote briefly from a draft document that won't be finalized until this autumn. It is sort of the Dutch equivalent of the Criteria Document for particulate matter, the Netherlands Aerosol Program. On particulate, it writes:

From the standpoint of dose, there appears to be little coherence between the epidemiologic and toxicologic studies. While the former show association of increased mortality and morbidity with acute exposure to PM at ambient concentrations below the current standards, the latter show associations of biological responses with PM atmosphere, both concentrated ambient PM and PM surrogates, only at orders of magnitude higher than ambient levels. A number of toxicologic studies with concentrated ambient particulate matter have shown no obvious relationship between exposure concentration and response.

Now, this is old news to many people in the room here, I am sure, but it seems to me also still startling news. I mean, if there is no obvious association between exposure and response, then what is the basis for presuming that reducing exposures slightly by enforcing NAAQS for PM is

going to, in fact, improve response?

Let's look, if we may, for just a moment on what I think the concentrated ambient particle studies in the lab have shown so far. I think we would all agree that CAPs studies are a tremendous improvement in this area, and I have the greatest respect and, in fact, awe for people who try to do this work, because this is phenomenally difficult work in the laboratory.

So far, with one exception, a report that was presented at a meeting in 1996 but could not be replicated by the investigators themselves and was never published in the peer reviewed literature, with that one exception, so far, CAPs, or concentrated air particles, don't kill animals. That is sort of unfortunate if you are a toxicologist. You know, if you are a toxicologist, and you believe that ambient PM kills people, concentrated ambient PM ought to kill lab animals. So far, that has not been possible to show, and that is sort of disappointing if you are an experimentalist.

Second, none of the inhaled CAP exposures that, at least, were reviewed in the draft Criteria Document...and, of course, there aren't that many studies, because this is still very hard stuff to do...none of those studies appear to have seriously affected either healthy or even compromised laboratory animals, various rodent models, a dog model, and many of the slight effects that have been observed appear to be reversible within about a day.

Very importantly, I think, some of the noted

changes such as recruitment of neutrophils are, of course, the normal and appropriate response of the immune system.

And, I think, much more importantly, the late Professor Glenn Cass and his colleagues, Dr. Ann McGill and others, have shown for years now that one of the very important fractions of PM, PM<sub>10</sub>, is biological. In particular, if you look at protein levels in air, even in Los Angeles, not a place known for the natural world anymore, even in Los Angeles, PM<sub>10</sub> is about...ambient air contains about 1 to 6 Fg/m<sup>3</sup> of total extractable protein. All of that is in the PM<sub>10</sub> fraction, by the way, and some of it is in the PM<sub>2.5</sub> fraction.

Now, ask yourself the question, if you are concentrating air particles off of Huntington Avenue in Boston by 30 fold, you are also, by definition, concentrating whatever proteins, lipopolysaccharides, and other important macromolecules are in that fine fraction. Now, some, of course, will be in the coarse fraction, but some are in the fine fraction, as has been shown by many people.

And, by the way, virtually none of this literature is reviewed in the Criteria Document, unfortunately, and I have provided, electronically to EPA by email and, I think, to the committee by email, a list of specific references in the peer reviewed literature that speak to this issue that I think should be included in the Criteria Document.

But ask yourself the question, if you are

concentrating ambient particles 30 fold and you are also concentrating all of these antigens and macromolecules 30 fold and you see an inflammatory response, what do you think it is due to? Right?

Yet, when you look in the laboratory, it saddens me to see that the CAPs investigators, to a man...and woman, I suppose...to a man, only analyze for M. You see these reports of the analysis of CAPs, and what do you see? You see organic carbon. That is supposed to be a representation of, if you think about it, probably thousands of different things, proteins, lipopolysaccharides, semi-volatile organic compounds, man-made stuff, natural stuff, antigens, non-antigens. It is very sad to me that Professor Cass' thoughts haven't really sort of translated into this CAPs field, at least yet, unless I am missing something.

So, I would suggest that when the Chapter 7 of the Criteria Document is reviewing the CAPs study and repeatedly noting information, there should be at least some discussion as to what that might mean, why you might get information from CAPs.

Okay, finally, as was talked about much more eloquently and knowledgeably by Dr. Vendetti who, I guess, had to leave, Chapter 7 reviews responses of various animal models of cardiopulmonary disease and their responses to CAPs. It is very clear that these models are very, very difficult to work with. It also seems clear, from Dr. Vendetti's comments, that the responses seen to date either

don't have clinical significance or seem to be inconsistent or, in any event, do not provide what the document calls biological plausibility. I think that is a stretch, at best, and really does not represent what cardiologists or toxicologists would say about those data.

Okay, two more slides. I want to spend just a moment on the question of whether particulate matter causes chronic effects. It is startling to me that in a toxicology chapter, Chapter 7, that reviews some 300 studies, there are only 3 on long-term effects, 3, and none of those give an indication that moderate levels of PM are harmful over the long term.

There doesn't seem to be any attempt to ask whether the toxicology on chronic effects is supportive of observations or not.

With respect to observations epidemiologically, I think there are two things to say. There have been many things said, of course, about the Pope, et al studies and the update in the JAMA article, but I would just like to mention two. The first is that, as Professor Vedal and others have pointed out really many years ago now, to call these studies long-term exposure studies and evidence of long-term effects is really to put an interpretation on them that is really not justified.

These are, of course, between city studies. They are not necessarily long-term studies. They are not necessarily long-term effects, even though people live in

cities for a long term.

Second, I just want to highlight what has also been mentioned before but to give you some numbers. Nowhere in that JAMA 2002 Pope article can you find the fact that 64 percent of men in that cohort and 55 percent of women in that cohort have some education beyond high school. And for all of those men and women, there appears to be no PM effect at all. The best estimate of effect for lung cancer mortality, all-cause mortality, cardiopulmonary mortality, and all other cause mortality, all four mortality estimates, the odds ratio is 1.0 of the best estimate.

Now, of course, the confidence intervals overlap the confidence intervals for everything else. Nonetheless, I think it is probably fair to say that virtually everyone in this room has some education beyond high school, which means the Pope, et al study suggests that for everyone in this room, there is no PM effect at all.

Final slide, please. So, you know, it suggests that maybe we should be building a lot more community colleges.

I have four suggestions, please, on how I think Chapters 7 and 9 of the draft Criteria Document might be improved. First, frankly, Chapter 7 doesn't look like a toxicology chapter to me. If you look at another Federal agency, the ATSDR which, of course, has been creating toxicologic profiles for a long time, they provided lots of useful information, none of which is in Chapter 7.



And I don't mean to be hypercritical here. Whoever worked on Chapter 7 had a lot of work to do, and it was, obviously, an enormous amount of work. There are 300 studies. It is a lot to do.

Yet, what do you do as a toxicologist? Well, you look for NOELs and LOELs. There isn't a single mention. It's not even in the glossary. You can't even find these terms in the glossary of the CD.

ATSDR, for its tox profiles, always plots up NOELs and LOELs, differentiates effects as serious and less serious, provides you with pictures as well as tables so that you can see by eye, on a log scale, where the LOELs are, where the NOELs are, and where ambient levels are.

It is completely missing from Chapter 7. I wish that all the hard work that went into Chapter 7...a lot of it didn't have to be done. You didn't have to review all that stuff and write it down. You just had to think about it quantitatively and summarize it in a useful way, and, unfortunately, that hasn't been done, and I think it is a great shame, and I hope it can be done now. So, just open up an ATSDR tox profile and copy the format, and I think it will be a vastly better chapter. I know, easier said than done.

Second, I think there needs to be explicit discussion in Chapter 9 as to whether the toxicologic data presented in Chapter 7 do or do not provide direct evidence for specific morbidity and mortality associations seen in the

observational epidemiologic literature.

Please do not continue to use the phrase biological plausibility. I mean, that is just hooey. Of course, it is biologically plausible. I mean, anything in air that is not either inert or oxygen biologically plausibly is bad for you. I mean, duh, you don't need toxicology for that.

What you need toxicology for is, to an experimental situation, expose animals or human volunteers over the short term to graded exposures to PM of various kinds and see what happens. That is what you need toxicology to do, not to provide biological plausibility.

There was a lot of biological plausibility that hormone replacement therapy could save women from heart disease. Bad news on that one.

Okay, finally, I think there needs to be much more emphasis on even the very best epidemiology studies, the case crossover studies that Dr. Schwartz spoke about before. Those are a tremendous improvement in design. But they are still missing a tremendous number of things.

If you look more broadly on the epidemiology on myocardial infarction and why MI rates vary from day to day, you find a whole wealth of things that are not measured in these studies. In Stockholm, for example, there is a major Stockholm heart epidemiology study, Dr. Jeda Mola and her colleagues, and what she finds...it is what your mother always told your father when they are arguing, you know, be quiet, Sam, you are going to give yourself a coronary.

Right? She finds that anger is a tremendous important risk factor for myocardial infarction, 15-fold elevated risk for MI if you have experienced extreme anger within one hour before that MI, 15 fold.

Now, if these case crossover studies that Professor Schwartz and others conduct...and they are incredibly elegant. I am awed by the statistical power in this room, frankly, and I teach at MIT. This is an impressive group of statisticians here working for the Agency and with the Agency. But how are you going to control for confounders you haven't measured?

If, in doing these studies, you don't know who is angry, you don't know who is anxious, you don't know who is stressed, and you think there might be some correlations between those things and fluctuations in PM, then all the models in the world can't turn bad apples into good applesauce.

I am sorry I have gone too long. I will stop.  
Thank you.

**DR. HOPKE:** Okay, thank you. Petros?

**DR. KOUTRAKIS:** Can I say something here?  
First, the concentrator concentrates particles...

**SPEAKER:** They can't hear you, Petros.

**DR. KOUTRAKIS:** The concentrator concentrates particles below 1.2 or something like that down to 0.1. Most of the biological material is in the coarse

fraction, and...

**DR. GREEN:** That is not true.

**DR. KOUTRAKIS:** Well, that is your opinion, and this is my opinion.

**DR. GREEN:** No, no, no. I provide...

**DR. MCCLELLAN:** Petros, has it been measured?

**DR. KOUTRAKIS:** Yes. At the beginning, we started measuring, and, also, we measured endotoxin which is more important, and we never found endotoxin. That is one question, clarification.

Next, as a scientist now and also with CASAC, I am kind of tired of coming here or elsewhere and have these smart consultants, they understand everything, and the bunch of us, we have no clue what we are doing. So, I would suggest that all these consultants go and apply for a grant to NIHS or EPA and do the research and explain to us what is happening, because in the laboratory, repeatedly, we can reproduce health effects using concentrated particles. Okay? Using ROFA.

**SPEAKER:** I'm sorry, did you say can or cannot?

**DR. KOUTRAKIS:** We can, we can. Okay? So, we seem to be very confused, so we will appreciate, you know, you to participate and rule out this hypothesis, because we do find effects all the time.

**DR. GREEN:** Can I say something or no?

**DR. HOPKE:** Quickly.

**DR. GREEN:** Okay. First, I guess I am getting the brunt for you consultants, and that is the problem with being last.

There are a lot of published data on lipopolysaccharide and protein content of PM<sub>5</sub> and lower, so endotoxin is one thing, but there is a whole world of macromolecules.

Second, I meant what I said, that I have the greatest respect, and I mean that, for everyone working in this field. What I think is important, though, is that what people are doing in this field, I think, is finding effects maybe for some fractions of PM and not others. Maybe the reason there is no dose-response yet is that people haven't looked at macromolecules, they haven't looked at antigens, they haven't done what Dr. Ann McGill has done, for example, what she was doing with Glenn Cass on road dust in L.A., and maybe, if people started looking more broadly at biochemicals, for example, at things that recruit cytokines...I mean, the neutrophils, you know, the immune system was around a long time before the industrial revolution.

My hope, which I think is the same as yours...I hope it is the same as yours...is to try to understand what causal fractions of air pollution...you know, what about air

pollution is worth controlling, and I fear that if we keep pretending that we understand this better than we do and that because ROFA, for example, which is full of metals, causes effect, therefore, all PM causes effects...

**DR. KOUTRAKIS:** I did not say that.

**DR. GREEN:** No, but that...

**DR. KOUTRAKIS:** That is not the argument for ROFA.

**DR. HOPKE:** Well, anyway, I think we are getting far afield here. So, I think, let's call it a day...

**DR. LIOY:** You make a point about the issue of education. I am not sure where you were going with that.

One of the things that I worry about is the fact that environmental justice and environmental equity are two very major concerns in this country, and what it seems to me is that this study is saying that not as a modifier, but maybe this is where the populations at risk are, and maybe we should looking for attention to exposure and also health effects studies on that population.

I just am not sure what you were driving at.

**DR. GREEN:** There are two various ways to think about the education thing, very different. The first is the one you are implying. The second, let me tell you what I was implying. The problem with the Pope, et al study...and it is going to get worse as more and more

follow-up goes...is, as you may know, the risk factor data were gathered only once by one four-page questionnaire in 1982, and now, there are up to 16 years of follow-up, and there is even more, I suppose, ongoing.

Now, ask yourself the question, which group of people is more likely to quit smoking? Focus now on the people in 1982. You have a bunch of Ph.D.s who are still smoking cigarettes, and a bunch of people who didn't complete the 9th grade smoking cigarettes. Up to 16 years pass. Some of them die. Which group is more likely to contain former smokers who, in the 1980s and 1990s, give up smoking, the Ph.D.s or the high school dropouts?

Now, I fear that, because we don't have information about all these other risk factors, who develops diabetes over those 16 years, who develops high blood pressure not to mention smoking, that what is missing from the models and what Dr. Pope and Thurston and all the other very brilliant statisticians can't make up for, cannot make up for, is that missing information.

Now, your hypothesis which, I think, is also interesting and potentially correct, part or in whole, is well, the smart people stay inside around like this, and the people who only got eighth grade education are out working construction. So, they are the ones being exposed.

I have got to tell you when you look, for example, at New York City, all right, look at who runs the New York City Marathon, look at who walks to work, look at where the





**CAPTION**

The Meeting in the matter, on the date, and at the time and place set out on the title page hereof.

It was requested that the Meeting be taken by the reporter and that the same be reduced to typewritten form.